Joint Image Reconstruction and Motion Estimation for Spatiotemporal Imaging*

Chong Chen†, Barbara Gris‡, and Ozan Öktem§

Abstract. We propose a variational model for joint image reconstruction and motion estimation applicable to spatiotemporal imaging. This model consists of two parts, one that conducts image reconstruction in a static setting and another that estimates the motion by solving a sequence of coupled indirect image registration problems, each formulated within the large deformation diffeomorphic metric mapping framework. The proposed model is compared against alternative approaches (optical flow based model and diffeomorphic motion models). Next, we derive efficient algorithms for a time-discretized setting and show that the optimal solution of the time-discretized formulation is consistent with that of the time-continuous one. The complexity of the algorithm is characterized and we conclude by giving some numerical examples in 2D space + time tomography with very sparse and/or highly noisy data.

Key words. spatiotemporal imaging, image reconstruction, motion estimation, joint variational model, shape theory, large diffeomorphic deformations

AMS subject classifications. 65F22, 65R32, 65R30, 65D18, 65J22, 65J20, 65L09, 68U10, 94A12, 94A08, 92C55, 54C56, 57N25, 47A52

1. Introduction.

Image reconstruction is challenging in a spatiotemporal setting where the object being imaged undergoes a temporal evolution during the data aquisition. This arises in tomographic imaging of the heart or lungs [13, 29] where the image recovery needs to include estimating and compensating for the unknown motion of the organs. As an example, data in positron emission tomography (PET) cardiac imaging is acquired over a relatively long period of time (often in the range of minutes). Here, respiratory and cardiac motion displaces organs of interest with up to 20–40 mm [48, 58]. Failing to correct for such motion leads to a degradation in image quality [29].

Data in spatiotemporal imaging is time (or quasi-time) dependent and an obvious approach is to decompose it into sub-sets (gates) such that data within each gate can be assumed to originate from the object in a fixed temporal state [25, 35, 37]. As an example, in the context of PET imaging of the heart or lungs, the gates correspond to the phases of breathing and/or cardiac motion [23, 32, 15]. Clearly a too fine gating leads to data with worse signal-to-noise-ratio, whereas to coarse gating leads to motion artefacts. Much work has been done on how to optimize the gating in order to obtain the optimal balance between signal-to-noise-ratio and motion artefacts [27, 56].

Algorithms for spatiotemporal image reconstruction that act on gated data either perform image reconstruction followed by motion estimation (sequential approach) or perform these

*Submitted to the editors 2018-12-18.

Funding: The work of all authors was supported by the Swedish Foundation for Strategic Research grant No. AM13-0049. Chen was in addition supported by the Beijing Natural Science Foundation grant No. Z180002.

†LSEC, ICMSEC, Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing 100190, China (chench@lsec.cc.ac.cn).

‡LJLL–Laboratoire Jacques-Louis Lions, Sorbonne Université, 75005 Paris, France (bgris.maths@gmail.com).

§Department of Mathematics, KTH–Royal Institute of Technology, 100 44 Stockholm, Sweden (ozan@kth.se).

This manuscript is for review purposes only.

two steps jointly [7, 29]. In the sequential approach, one starts with applying static image reconstruction on data from each of the gates, resulting in a series of (low-resolution) images. Then, one selects a reconstructed image as target and registers the other reconstructed images against this target. The final reconstructed image is obtained as an average of all the registered images [2, 3, 22, 24, 30, 29]. The other alternative that seeks to jointly perform image reconstruction and motion estimation is much more complex. Several approaches have been suggested for how to do this, such as [7, 47, 51, 28, 36, 9, 8, 44, 34, 52, 6, 33, 33, 10, 13].

The approach taken in this paper seeks to jointly perform image reconstruction and motion estimation. The motion is here modelled as deformations induced by diffeomorphisms that are provided by the large deformation diffeomorphic metric mapping (LDDMM) framework, which is a well-developed framework for diffeomorphic image registration [54, 26, 38, 5, 31, 60, 55, 11, 12]. Joint image reconstruction and motion estimation can be based on the growth model in LDDMM [31] as shown in [33] for 3D space + time computed tomography (CT). It can also be used for indirect image registration as shown in [17], which also proves the resulting model is stable.

Specific contributions. The main contribution is a new variational model for joint image reconstruction and motion estimation in spatiotemporal imaging based on the LDDMM framework with deformable templates. As already mentioned, there are two components: one corresponding to modified static image reconstruction, and the other to estimating the motion. To handle the latter, we generalize the LDDMM framework to a setting that applies to a series of coupled sequential indirect image registration problems.

The mathematical properties of the proposed variational model is compared against the optical flow based model in [13] and the diffeomorphic motion model in [33]. The comparison shows that the proposed model has some desirable properties in terms of the optimal solution, e.g., guaranteeing elastically large diffeomorphic deformations, averagely distributed w.r.t. time t, and non-vanishing at both the initial and the end time points, etc. Moreover, a computationally efficient gradient-based iterative scheme is presented for a time-discretized formulation. More importantly, the optimal solution of the time-discretized problem is shown to be consistent with that of the time-continuous one. Most of the computationally demanding parts relate to computations involving linearized deformations [43].

Outline. Section 2 presents a general variational model for joint image reconstruction and motion estimation. We also briefly review necessary parts of the LDDMM theory in subsection 3.1. Subsection 3.2 proposes the new variational model and subsection 3.3 makes the theoretical comparison between this ans other existing models. Section 4 gives the detailed numerical algorithms associated with the proposed model. Results of numerical experiments are presented in section 5 in the context of 2D space + time tomography. Finally, we conclude with section 6 that discusses future work.

2. A variational model for joint image reconstruction and motion estimation.

Spatiotemporal imaging often leads to the task of recovering a spatially distributed quantity (image) that exhibits temporal variations given indirect time-dependent noisy observations (measured data). Hence, both the image and its motion are unknown.

2.1. General spatiotemporal inverse problem.

To formalise the notion of spatiotemporal imaging, let $f\colon [t_0, t_1] \times \Omega \to \mathbb{R}^k$ denote the spatiotemporal image we need to recover. Here k

is the number of channels/modalities (often $k = 1$) and $\Omega \subset \mathbb{R}^n$ is the spatial region containing the support of the image. Without loss of generality, the general (quasi-)time domain $[t_0, t_1]$ can be re-parameterized onto $[0, 1]$.

The spatiotemporal inverse problem is now stated mathematically as the problem of reconstructing the spatiotemporal image $f(t, \cdot) \in \mathscr{X}$ given measured data $g(t, \cdot) \in \mathscr{Y}$ where

$$g(t, \cdot) = \mathcal{T}\big(t, f(t, \cdot)\big) + g_{\text{noise}}(t, \cdot) \quad \text{for } t \in [0, 1]. \tag{2.1}$$

Here, $\mathscr{X}$ (reconstruction space) is the vector space of all possible images on a fixed domain Ω, $\mathscr{Y}$ (data space) is the vector space of all possible data, and $g_{\text{noise}}(t, \cdot) \in \mathscr{Y}$ is the observation noise in data. Furthermore, $\mathcal{T}(t, \cdot)\colon \mathscr{X} \to \mathscr{Y}$ is a time-dependent forward operator, for short denoted by $\mathcal{T}_t$, that models how an image at time t gives rise to data in absence of noise or measurement errors (e.g., a stack of ray transforms with time varying data acquisition parameters in CT or the forward model in PET with time dependent attenuation, etc.) [41].

To proceed, we further specify the form of the spatiotemporal image $f(t, \cdot)$ by making use of deformable templates from shape theory.

2.2. Spatiotemporal inverse problem and shape theory. Shape theory seeks to develop quantitative tools to study shapes and their variability. The theory can be traced back to work by D'Arcy Thompson [53]. The underlying idea is that shapes of objects are represented as a deformation that acts on a template [55]. Hence, the template is the "shape invariant" part of the object whereas the set of deformations model the various shapes that can be generated from the template. Shape similarity between two objects can then be quantified as the "cost" of deforming one object into the other by means of a minimal deformation in the set of deformations. Further details are given in [60, 31, 40].

Bearing in mind the above, we separate the spatial and temporal components of a spatiotemporal image as

$$f(t, \cdot) := \mathcal{W}(\phi_t, I) \quad \text{for some } \phi_t \in \mathscr{G} \text{ and } I \in \mathscr{X}. \tag{2.2}$$

Here $I\colon \Omega \to \mathbb{R}$ (template) is the time-independent spatial component, $\mathcal{W}\colon \mathscr{G} \times \mathscr{X} \to \mathscr{X}$ is a deformation operator the models how a deformation parameter $\phi_t \in \mathscr{G}$ deforms the an image, and $\phi_t\colon \Omega \to \Omega$ is the temporal evolution of the deformation parameter.

We will assume deformations are given by diffeomorphic group actions, so $\mathscr{G}$ is a suitable subgroup of the diffeomorphic group on Ω and $\mathcal{W}$ is given by the group action of $\mathscr{G}$ on $\mathscr{X}$, i.e., $\mathcal{W}(\phi_t, I) := \phi_t \,.\, I$. The spatiotemporal inverse problem can then be written as

$$g(t, \cdot) = \mathcal{T}_t(\phi_t \,.\, I) + g_{\text{noise}}(t, \cdot) \quad \text{for } t \in [0, 1]. \tag{2.3}$$

Notice that $f(t, \cdot) = \phi_t \,.\, I$ is the spatiotemporal image at time t generated from the template I by the diffeomorphism ϕ_t. Hence, the above inverse problem calls for jointly recovering the (time-independent) template I and the curve of diffeomorphisms $t \mapsto \phi_t$.

Common group actions. In imaging, there are several diffeomorphic group actions that one could consider. A natural one merely moves intensities without changing them (geometric deformations) [17]:

$$\phi_t \,.\, I := I \circ \phi_t^{-1}. \tag{2.4}$$

An alternative group action corresponds to mass-preserving deformations [17, 60]:

$$\phi_t \,.\, I := \left|D(\phi_t^{-1})\right| I \circ \phi_t^{-1}. \tag{2.5}$$

Here, $\left|D(\phi)\right|$ denotes the determinant of the Jacobian of ϕ. This group action adjusts the intensity values but preserves the total mass.

2.3. Joint image reconstruction and motion estimation. It is not difficult to observe that the inverse problem in (2.3) decomposes into two sub-problems. Given a diffeomorphism ϕ_t, the original problem reduces to a static image reconstruction problem where one seeks to recover the template "I" from noisy measured data. The data represents indirect noisy measurements of the template that has undergone a known diffeomorphic deformation. Conversely, given the template I, the original problem reduces to an indirect registration problem at each point in time t. More precisely, the task is to recover a curve of diffeomorphisms "$t \mapsto \phi_t$" that registers the template at time t against a target that is known indirectly through noisy measured data $g(t, \cdot)$. Such a "sequential indirect image registration problem" can be seen as a temporal version of the indirect image registration problem in [17].

Inverse problem of the type (2.3) in imaging applications are often ill-posed. A flexible framework for regularizing many of these problems is through a variational formulation [46]. The idea is to add regularization functional that penalizes a maximum likelihood solution and thereby acts as a stabiliser by preventing overfitting. The variational formulation for (2.3) reads as

$$\min_{\substack{I\in\mathscr{X}\\ \phi_t\in\mathscr{G}}} \left\{ \int_0^1 \left[\mathcal{D}\big(\mathcal{T}_t(\phi_t \,.\, I), g(t,\cdot)\big) + \mu_2 \mathcal{R}_2(\phi_t) \right] \mathrm{d}t + \mu_1 \mathcal{R}_1(I) \right\} \quad \text{for fixed } \mu_1, \mu_2 > 0. \tag{2.6}$$

In the above, μ_1, μ_2 are the regularization parameters that must be chosen depending on the noise level in data. Next, $\mathcal{D}\colon \mathscr{Y} \times \mathscr{Y} \to \mathbb{R}_+$ above is the data discrepancy functional that quantifies the mismatch in data space $\mathscr{Y}$. It is often taken as a suitable affine transform of the negative data likelihood for the data, so minimising it amounts to computing a maximim likelihood estimator. If one has data with additive Gaussian noise, then it is given by the squared $\mathscr{L}^2$-norm:

$$\mathcal{D}\big(g_1, g_2\big) := \big\|g_1 - g_2\big\|_2^2 \quad \text{for } g_1, g_2 \in \mathscr{Y}. \tag{2.7}$$

Likewise, Poisson distributed data leads to the Kullback-Leibler (KL) divergence:

$$\mathcal{D}\big(g_1, g_2\big) := \int_{\mathbb{M}} g_1(y) - g_2(y) \log\big(g_1(y)\big)\, \mathrm{d}y \quad \text{for } g_1, g_2 \in \mathscr{Y}. \tag{2.8}$$

Moreover, the spatial regularization $\mathcal{R}_1\colon \mathscr{X} \to \mathbb{R}_+$ introduces stability by encoding priori knowledge about the ground truth image. It is frequently based on a roughness prior given as an $\mathscr{L}^p$-norm of the gradient magnitude or $\mathscr{L}^1$-norm of a suitable sparse representation. Typically, if $\mathscr{X} := \mathcal{H}^1(\Omega)$ then one takes the squared $\mathscr{L}^2$-norm of the gradient magnitude:

$$\mathcal{R}_1(f) := \|\nabla f\|_2^2 \tag{2.9}$$

This choice is known to produce smooth images whereas selecting the $\mathscr{L}^1$-norm of the gradient magnitude for $\mathscr{X} := \mathscr{BV}(\Omega)$ is better at preserving edges (total variation (TV) regularization) [45]:

$$\mathcal{R}_1(f) := \|\nabla f\|_1 \tag{2.10}$$

What remains is to describe how to generate the curve $t \mapsto \phi_t$ and to select the shape regularization $\mathcal{R}_2 \colon \mathscr{G} \to \mathbb{R}_+$. An overall difficulty is that diffeomorphisms $\mathscr{G}$ do not form a vector space. One option is to try re-parametrising the deformations using vector space elements. As we shall see next, this can be done within the LDDMM framework.

3. A variational model for joint image reconstruction and motion estimation.

This section introduces a new variational model of the framework (2.6) based on LDDMM. We begin by recalling the basic principles of LDDMM.

3.1. The LDDMM framework.

The LDDMM framework outlined here offers a powerful machinery for generating a flow of diffeomorphisms by means of a velocity field. One advantage is that the set of velocity fields form a vector space, so we have in this way re-parametrised deformations by vector space elements.

The idea in LDDMM is to consider a sequence of infinitesimally small vector field perturbations to the identity mapping. These vector fields can be seen as an instantaneous velocity field. Under certain regularity, the composition of such small deformations in the limit generates a flow of diffeomorphisms given as the solution to an ordinary differential equation (ODE) [60]. More precisely, given a velocity field $\boldsymbol{\nu} \colon [0,1] \times \Omega \to \mathbb{R}^n$, a flow $t \mapsto \phi_t$ is generated by the following ODE:

$$\begin{cases} \partial_t \phi_t(x) = \boldsymbol{\nu}\big(t, \phi_t(x)\big) \\ \phi_0(x) = x \end{cases} \qquad \text{for } x \in \Omega \text{ and } 0 \le t \le 1. \tag{3.1}$$

Note that $\phi_0 = \mathrm{Id}$, i.e., the flow starts at the identity mapping. If the velocity field $\boldsymbol{\nu}$ is sufficiently regular (see Definition 3.1), then the above ODE has a well-defined solution that is a diffeomorphism at each time point, i.e., (3.1) defines a flow of diffeomorphisms.

Definition 3.1 (Admissible space [60]). *A Hilbert space $\mathscr{V} \subset \mathscr{C}^1_0(\Omega, \mathbb{R}^n)$ is admissible if it is (canonically) embedded in $\mathscr{C}^1_0(\Omega, \mathbb{R}^n)$ with the $\|\cdot\|_{1,\infty}$ norm, i.e., there exists a constant $C > 0$ such that*

$$\|\boldsymbol{v}\|_{1,\infty} \le C \|\boldsymbol{v}\|_{\mathscr{V}} \quad \text{for all } \boldsymbol{v} \in \mathscr{V}.$$

In the above, $\|\boldsymbol{v}\|_{1,\infty} := \|\boldsymbol{v}\|_\infty + \|D\boldsymbol{v}\|_\infty$ for $\boldsymbol{v} \in \mathscr{C}^1_0(\Omega, \mathbb{R}^n)$.

We now consider velocity fields that are $\mathscr{L}^p$ in time and in an admissible space at every point in time, i.e.,

$$\mathscr{L}^p([0,1], \mathscr{V}) := \Big\{ \boldsymbol{\nu} \colon \boldsymbol{\nu}(t, \cdot) \in \mathscr{V} \text{ and } \|\boldsymbol{\nu}\|_{\mathscr{L}^p([0,1],\mathscr{V})} < \infty \text{ for } 1 \le p \le \infty \Big\} \tag{3.2}$$

with the associated norm

$$\|\boldsymbol{\nu}\|_{\mathscr{L}^p([0,1],\mathscr{V})} := \left(\int_0^1 \big\|\boldsymbol{\nu}(t, \cdot)\big\|_{\mathscr{V}}^p \, \mathrm{d}t \right)^{1/p}.$$

To simplify notation, let $\mathscr{L}^p_{\mathscr{V}}(\Omega)$ denote $\mathscr{L}^p([0,1],\mathscr{V})$ and note in particular that $\mathscr{L}^2_{\mathscr{V}}(\Omega)$ is a Hilbert space with inner product

$$\langle \boldsymbol{\nu}, \boldsymbol{\eta} \rangle_{\mathscr{L}^2_{\mathscr{V}}(\Omega)} = \int_0^1 \big\langle \boldsymbol{\nu}(t,\cdot), \boldsymbol{\eta}(t,\cdot) \big\rangle_{\mathscr{V}} \,\mathrm{d}t \quad \text{for } \boldsymbol{\nu}, \boldsymbol{\eta} \in \mathscr{L}^2_{\mathscr{V}}(\Omega).$$

The following theorem states that admissible velocity field generates a flow of diffeomorphisms.

Theorem 3.2 ([60, 11]). *Let $\mathscr{V}$ be an admissible Hilbert space and $\boldsymbol{\nu} \in \mathscr{L}^2_{\mathscr{V}}(\Omega)$ be a velocity field. Then the ODE in (3.1) admits a unique solution $\phi^{\boldsymbol{\nu}} \in \mathscr{C}^1_0([0,1]\times\Omega,\Omega)$, such that for $t \in [0,1]$, the mapping $\phi^{\boldsymbol{\nu}}_t \colon \Omega \to \Omega$ is a $\mathscr{C}^1$-diffeomorphism on Ω.*

Hence, we define

$$\mathscr{G}_{\mathscr{V}} := \Big\{ \phi \colon \phi = \phi^{\boldsymbol{\nu}}_{0,1} \text{ for some } \boldsymbol{\nu} \in \mathscr{L}^2_{\mathscr{V}}(\Omega) \Big\}, \tag{3.3}$$

which becomes a sub-group of the group of diffeomorphisms by Theorem 3.2 whenever $\mathscr{V}$ is admissible. Next, if $\phi^{\boldsymbol{\nu}}_t$ denotes the solution to the ODE in (3.1) with given $\boldsymbol{\nu} \in \mathscr{L}^2_{\mathscr{V}}(\Omega)$, then

$$\phi^{\boldsymbol{\nu}}_{s,t} := \phi^{\boldsymbol{\nu}}_t \circ (\phi^{\boldsymbol{\nu}}_s)^{-1} \quad \text{for } 0 \le t, s \le 1 \tag{3.4}$$

Furthermore, $\phi^{\boldsymbol{\nu}}_0 = \mathrm{Id}$ and by (3.4) we also get

$$\phi^{\boldsymbol{\nu}}_t = \phi^{\boldsymbol{\nu}}_{0,t}, \quad (\phi^{\boldsymbol{\nu}}_t)^{-1} = \phi^{\boldsymbol{\nu}}_{t,0}. \tag{3.5}$$

Next, several important properties about $\mathscr{G}_{\mathscr{V}}$ are summarised in the following theorem.

Theorem 3.3 ([60, 11]). *Let $\mathscr{V}$ be an admissible Hilbert space, $\mathscr{G}_{\mathscr{V}}$ is defined as in (3.3), and $\mathrm{d}_{\mathscr{G}_{\mathscr{V}}} : \mathscr{G}_{\mathscr{V}} \times \mathscr{G}_{\mathscr{V}} \to \mathbb{R}_+$ is defined as*

$$\mathrm{d}_{\mathscr{G}_{\mathscr{V}}}(\phi,\psi) := \inf_{\substack{\boldsymbol{\nu}\in\mathscr{L}^2_{\mathscr{V}}(\Omega) \\ \psi = \phi\circ\phi^{\boldsymbol{\nu}}_{0,1}}} \|\boldsymbol{\nu}\|_{\mathscr{L}^2_{\mathscr{V}}(\Omega)} \quad \textit{for } \phi,\psi \in \mathscr{G}_{\mathscr{V}}. \tag{3.6}$$

Then $\mathscr{G}_{\mathscr{V}}$ is a group for the composition of functions and $\mathscr{G}_{\mathscr{V}}$ is a complete metric space under the metric $\mathrm{d}_{\mathscr{G}_{\mathscr{V}}}$. Furthermore, for each $\phi,\psi \in \mathscr{G}_{\mathscr{V}}$, there exists $\boldsymbol{\nu} \in \mathscr{L}^2_{\mathscr{V}}(\Omega)$ satisfying $\psi = \phi \circ \phi^{\boldsymbol{\nu}}_{0,1}$, i.e., $\mathrm{d}_{\mathscr{G}_{\mathscr{V}}}(\phi,\psi) = \|\boldsymbol{\nu}\|_{\mathscr{L}^2_{\mathscr{V}}(\Omega)}$.

The metric $\mathrm{d}_{\mathscr{G}_{\mathscr{V}}}$ can now be used to define an energy term for regularizing the registration of a template image I_0 to a target image I_1:

$$\inf_{\phi\in\mathscr{G}_{\mathscr{V}}} \Big\{ \big\| \phi \,.\, I_0 - I_1 \big\|^2_{\mathscr{L}^2(\Omega)} + \mu\, \mathrm{d}_{\mathscr{G}_{\mathscr{V}}}(\mathrm{Id},\phi)^2 \Big\} \quad \text{for fix regularization parameter } \mu > 0. \tag{3.7}$$

One can show that the infimum in (3.6) is reached, so it can be replaced by a minimum. Next one can also show [60, Lemma 11.3] that (3.7) is equivalent to the following variational model that optimizes over velocity fields instead of diffeomorphisms:

$$\min_{\boldsymbol{\nu}\in\mathscr{L}^2_{\mathscr{V}}(\Omega)} \Big\{ \big\| \phi^{\boldsymbol{\nu}}_{0,1}.I_0 - I_1 \big\|^2_{\mathscr{L}^2(\Omega)} + \mu \int_0^1 \big\| \boldsymbol{\nu}(t,\cdot) \big\|^2_{\mathscr{V}} \,\mathrm{d}t \Big\} \quad \text{with } \phi^{\boldsymbol{\nu}}_{0,1} \in \mathscr{G}_{\mathscr{V}} \text{ as in (3.4)}. \tag{3.8}$$

In conclusion, the regularization term for image registration by LDDMM is

$$\mathcal{R}(\phi) := \mathrm{d}_{\mathscr{G}_\mathscr{V}}(\mathrm{Id}, \phi)^2 = \int_0^1 \big\|\widehat{\boldsymbol{\nu}}(t,\cdot)\big\|_\mathscr{V}^2 \,\mathrm{d}t, \tag{3.9}$$

where $\widehat{\boldsymbol{\nu}}$ above minimizes $\boldsymbol{\nu} \mapsto \mathrm{d}_{\mathscr{G}_\mathscr{V}}(\mathrm{Id}, \phi) = \|\boldsymbol{\nu}\|_{\mathscr{L}^2_\mathscr{V}(\Omega)}$ where $\phi = \mathrm{Id} \circ \phi^{\boldsymbol{\nu}}_{0,1}$ (such a minimiser exists due to Theorem 3.3).

A final remark concerns the choice of $\mathscr{V}$. If we choose it as a reproducing kernel Hilbert space (RKHS) with a symmetric and positive-definite reproducing kernel, then $\mathscr{V}$ becomes an admissible Hilbert space [11]. Using such vector fields is also advantageous from a computational point of view as shown in [17]. We will therefore assume that $\mathscr{V}$ is henceforth an admissible RKHS.

3.2. Spatiotemporal reconstruction with LDDMM. The aim here is to formulate a special case of (2.6) for solving the spatiotemporal inverse problem in (2.3). Let ϕ_t in (2.6) be generated by the flow equation (3.1) as in LDDMM, i.e., $\phi_t = \phi^{\boldsymbol{\nu}}_{0,t}$ for some velocity field $\boldsymbol{\nu}$. If $\boldsymbol{\nu} \in \mathscr{L}^2_\mathscr{V}(\Omega)$, then $\phi^{\boldsymbol{\nu}}_t$ is a diffeomorphism on Ω by Theorem 3.2. Consequently, combining Theorem 3.3 with (3.9) implies that a shape regularizer $\mathcal{R}_2$ for the temporal deformation given by $\phi^{\boldsymbol{\nu}}_{0,t}$ in (2.6) can be defined as

$$\mathcal{R}_2(\phi^{\boldsymbol{\nu}}_{0,t}) := \int_0^t \big\|\boldsymbol{\nu}(\tau,\cdot)\big\|_\mathscr{V}^2 \,\mathrm{d}\tau. \tag{3.10}$$

The variational formulation in (2.6) for solving the spatiotemporal inverse problem in (2.3) now reads as

$$\begin{aligned} &\min_{\substack{I\in\mathscr{X}\\ \boldsymbol{\nu}\in\mathscr{L}^2_\mathscr{V}(\Omega)}} \int_0^1 \left[\mathcal{D}\big(\mathcal{T}_t(\phi^{\boldsymbol{\nu}}_{0,t}\,.\,I), g(t,\,\cdot\,)\big) + \mu_2 \int_0^t \big\|\boldsymbol{\nu}(\tau,\cdot)\big\|_\mathscr{V}^2 \,\mathrm{d}\tau\right] \mathrm{d}t + \mu_1 \mathcal{R}_1(I) \\ &\qquad \text{s.t. } \phi^{\boldsymbol{\nu}}_{0,t} \text{ solves (3.1).} \end{aligned} \tag{3.11}$$

Note that (3.11) is an ODE constrained optimization problem, which henceforth is referred to as *time-continuous spatiotemporal reconstruction with LDDMM*.

Remark 3.4. Changing the order of integration for the second term in (3.11) yields the following equivalent formulation:

$$\begin{aligned} &\min_{\substack{I\in\mathscr{X}\\ \boldsymbol{\nu}\in\mathscr{L}^2_\mathscr{V}(\Omega)}} \int_0^1 \left[\mathcal{D}\big(\mathcal{T}_t(\phi^{\boldsymbol{\nu}}_{0,t}\,.\,I), g(t,\,\cdot\,)\big) + \mu_2(1-t)\big\|\boldsymbol{\nu}(t,\cdot)\big\|_\mathscr{V}^2\right] \mathrm{d}t + \mu_1 \mathcal{R}_1(I) \\ &\qquad \text{s.t. } \phi^{\boldsymbol{\nu}}_{0,t} \text{ solves (3.1).} \end{aligned} \tag{3.12}$$

3.2.1. Basic properties of optimal velocity fields. The aim here is to characterize temporal behaviour of velocity fields that solve (3.11) when $\mathscr{V}$ is a RKHS. For this purpose, we introduce some notation that will simplify the expressions:

$$\mathcal{D}_{g_t}(f) := \mathcal{D}\Big(\mathcal{T}_t(f), g(t,\,\cdot\,)\Big) \quad \text{for } f \in \mathscr{X} \text{ with given } g(t,\,\cdot\,) \in \mathscr{Y}. \tag{3.13}$$

By Theorem A.4 (in Appendix A), the optimal velocity field in (3.11) satisfies

$$\boldsymbol{\nu}(t,\,\cdot\,) = \frac{1}{2\mu_2(1-t)} \int_t^1 \mathcal{K}\Big(\nabla(\phi^{\boldsymbol{\nu}}_{0,t}\,.\,I)\big|D(\phi^{\boldsymbol{\nu}}_{t,\tau})\big|\,\nabla\mathcal{D}_{g_\tau}(\phi^{\boldsymbol{\nu}}_{0,\tau}\,.\,I)(\phi^{\boldsymbol{\nu}}_{t,\tau})\Big)\,\mathrm{d}\tau \quad \text{for } 0 \le t < 1. \tag{3.14}$$

Here, $\mathcal{K}(\varphi) = \int_\Omega \mathsf{K}(\,\cdot\,,y)\varphi(y)\,\mathrm{d}y$, with $\mathsf{K}\colon \Omega\times\Omega \to \mathbb{M}^{n\times n}_+$ denoting the reproducing kernel if $\mathscr{V}$ is a RKHS.

By the above, the optimal velocity field can be seen as a time-averaging. Next, $\boldsymbol{\nu}(1,\,\cdot\,)$ is well-defined at $t=1$ and reads as

$$\boldsymbol{\nu}(1,\cdot) = \frac{1}{2\mu_2}\mathcal{K}\big(\nabla(\phi^{\boldsymbol{\nu}}_{0,1}\,.\,I)\,\nabla\mathcal{D}_{g_1}(\phi^{\boldsymbol{\nu}}_{0,1}\,.\,I)\big). \tag{3.15}$$

Similarely, by (A.9) and (A.7) we get

$$\boldsymbol{\nu}(0,\cdot) = \frac{1}{2\mu_2}\mathcal{K}\bigg(\nabla I \int_0^1 \big|D(\phi^{\boldsymbol{\nu}}_{0,t})\big|\,\nabla\mathcal{D}_{g_t}(\phi^{\boldsymbol{\nu}}_{0,t}\,.\,I)(\phi^{\boldsymbol{\nu}}_{0,t})\,\mathrm{d}t\bigg). \tag{3.16}$$

Hence, the optimal velocity field in (3.11) is *averagely distributed w.r.t. time t*, and *non-vanishing at both the initial and the end time points.*

3.2.2. Control theoretic formulation. The aim here is to state an equivalent optimal control formulation of (3.11) in terms of a partial differential equation (PDE) constrained optimization problem. This makes it easier to compare our proposed approach in (3.11) against PDE based approaches for registration, as we do in subsection 3.3.

Theorem 3.5. *Let $\mathscr{X}$ be a space of real valued functions that are sufficiently smooth, e.g., the space of functions with bounded variation given in subsection 3.3.1. Assume next that $I \in \mathscr{X}$ and define $f\colon [0,1]\times\Omega\to\mathbb{R}$ as*

$$f(t,\,\cdot\,) := \phi^{\boldsymbol{\nu}}_{0,t}\,.\,I \quad \textit{for } 0\le t\le 1 \textit{ with } \phi^{\boldsymbol{\nu}}_{0,t} \textit{ given by (3.4).} \tag{3.17}$$

Assume furthermore that $f(t,\,\cdot\,)\in\mathscr{X}$ for any $0\le t\le 1$. Then, (3.11) is under the group action in (2.4) (geometric deformation) equivalent to

$$\begin{aligned}\min_{\substack{f(0,\,\cdot\,)\in\mathscr{X}\\ \boldsymbol{\nu}\in\mathscr{L}^2_{\mathscr{V}}(\Omega)}} &\left\{\int_0^1\left[\mathcal{D}\Big(\mathcal{T}_t\big(f(t,\,\cdot\,)\big),g(t,\,\cdot\,)\Big) + \mu_2\int_0^t \big\|\boldsymbol{\nu}(\tau,\cdot)\big\|^2_{\mathscr{V}}\,\mathrm{d}\tau\right]\mathrm{d}t + \mu_1\mathcal{R}_1\big(f(0,\,\cdot\,)\big)\right\}\\ &\textit{s.t. } \partial_t f(t,\,\cdot\,) + \big\langle\nabla f(t,\cdot),\boldsymbol{\nu}(t,\cdot)\big\rangle_{\mathbb{R}^n} = 0.\end{aligned} \tag{3.18}$$

Likewise, (3.11) is under the group action in (2.5) (mass-preserving deformation) equivalent to

$$\begin{aligned}\min_{\substack{f(0,\,\cdot\,)\in\mathscr{X}\\ \boldsymbol{\nu}\in\mathscr{L}^2_{\mathscr{V}}(\Omega)}} &\left\{\int_0^1\left[\mathcal{D}\Big(\mathcal{T}_t\big(f(t,\,\cdot\,)\big),g(t,\,\cdot\,)\Big) + \mu_2\int_0^t \big\|\boldsymbol{\nu}(\tau,\cdot)\big\|^2_{\mathscr{V}}\,\mathrm{d}\tau\right]\mathrm{d}t + \mu_1\mathcal{R}_1\big(f(0,\,\cdot\,)\big)\right\}\\ &\textit{s.t. } \partial_t f(t,\,\cdot\,) + \nabla\cdot\big(f(t,\,\cdot\,)\,\boldsymbol{\nu}(t,\,\cdot\,)\big) = 0.\end{aligned} \tag{3.19}$$

Proof. First we consider the geometric deformation in (2.4), so (3.17) reads as

$$f(t,\cdot) = I \circ (\phi^{\boldsymbol{\nu}}_{0,t})^{-1} \quad \text{for } 0 \le t \le 1. \tag{3.20}$$

Obviously, $f(0,\cdot) = I$, $f(1,\cdot) = I \circ (\phi^{\boldsymbol{\nu}}_{0,1})^{-1}$, and

$$f\big(t, \phi^{\boldsymbol{\nu}}_{0,t}\big) = I \quad \text{for } 0 \le t \le 1. \tag{3.21}$$

Differentiating (3.21) w.r.t. time t leads to

$$\partial_t f\big(t, \phi^{\boldsymbol{\nu}}_{0,t}\big) + \big\langle \nabla f\big(t, \phi^{\boldsymbol{\nu}}_{0,t}\big), \boldsymbol{\nu}\big(t, \phi^{\boldsymbol{\nu}}_{0,t}\big)\big\rangle_{\mathbb{R}^n} = 0.$$

The above is the PDE constraint in (3.18), so a solution to (3.11) also solves (3.18).

We now consider the reverse implication, i.e., show that a solution to (3.18) also solves (3.11). Suppose that f and $\boldsymbol{\nu}$ solve (3.18). Define the diffeomorphism ψ_t that solves the ODE in (3.1) with the above given $\boldsymbol{\nu}$. Since f satisfies PDE constraint in (3.18), we get

$$\frac{d}{dt} f(t,\psi_t) = \partial_t f(t,\psi_t) + \big\langle \nabla f(t,\psi_t), \boldsymbol{\nu}(t,\psi_t)\big\rangle_{\mathbb{R}^n} = 0.$$

Hence, $t \mapsto f\big(t,\psi_t\big)$ is constant and in particular we have

$$f\big(t,\psi_t\big) \equiv f\big(0,\psi_0\big) = f(0,\cdot).$$

If $f(0,\cdot) := I$ and $\psi_t := \phi^{\boldsymbol{\nu}}_{0,t}$, then $f\big(t,\cdot\big) = I \circ (\phi^{\boldsymbol{\nu}}_{0,t})^{-1}$. Hence a solution to (3.18) also solves (3.11).

Using a mass-preserving deformation (2.5) as group action in (3.17) results in

$$f(t,\cdot) = \big|D\big((\phi^{\boldsymbol{\nu}}_{0,t})^{-1}\big)\big| I \circ (\phi^{\boldsymbol{\nu}}_{0,t})^{-1} \quad \text{for } 0 \le t \le 1. \tag{3.22}$$

We then get that $f(0,\cdot) = I$ and $f(1,\cdot) = \big|D\big((\phi^{\boldsymbol{\nu}}_{0,1})^{-1}\big)\big|\, I \circ (\phi^{\boldsymbol{\nu}}_{0,1})^{-1}$. The symmetry of the mass-preserving property furthermore yields

$$\big|D(\phi^{\boldsymbol{\nu}}_{0,t})\big|\, f(t,\cdot) \circ \phi^{\boldsymbol{\nu}}_{0,t} = I \quad \text{for } 0 \le t \le 1. \tag{3.23}$$

Finally, differentiating (3.23) w.r.t. t leads to the constraint in (3.19). Hence, a minimizer of (3.11) with the group action given by (2.5) is also a minimizer of (3.19). Similar to the case of geometric deformation, it is not difficult to prove the reverse implication. ∎

3.3. Comparison with existing approaches. Here we compare (3.11) against several existing approaches (i.e., optical flow based model, diffeomorphic motion models).

3.3.1. Optical flow based models. Recently, [13] proposes an optical flow based variational model (joint TV-TV optical flow) for joint motion estimation and image reconstruction in spatiotemporal imaging. The model is formulated as a PDE-constrained optimal control problem where the constraint is given by a brightness constancy equation. When applied to the spatiotemporal inverse problem in (2.3), it reads as

$$\begin{aligned} \min_{\substack{f(t,\cdot)\in\mathscr{X}\\ \boldsymbol{\nu}(t,\cdot)\in\mathscr{BV}(\Omega)}} & \int_0^1 \Big[\mathcal{D}\Big(\mathcal{T}_t\big(f(t,\cdot)\big), g(t,\cdot)\Big) + \mu_1 \mathcal{R}_1\big(f(t,\cdot)\big) + \mu_2 \big|\boldsymbol{\nu}(t,\cdot)\big|_{\mathscr{BV}}\Big]\, \mathrm{d}t \\ \text{s.t. } & \partial_t f(t,\cdot) + \big\langle \nabla f(t,\cdot), \boldsymbol{\nu}(t,\cdot)\big\rangle_{\mathbb{R}^n} = 0, \end{aligned} \tag{3.24}$$

where $|\cdot|_{\mathscr{BV}}$ is the semi-norm on the space of functions with bounded variation:

$$\mathscr{BV}(\Omega) := \left\{ u \in \mathscr{L}^1(\Omega) : |u|_{\mathscr{BV}} < \infty \right\} \quad \text{with} \quad |u|_{\mathscr{BV}} := \sup_{\substack{\varphi \in \mathscr{C}_0^1(\Omega, \mathbb{R}^n) \\ \|\varphi\|_\infty \leq 1}} \int_\Omega u(x) \operatorname{div} \varphi(x) \, \mathrm{d}x.$$

In particular, $\left|\boldsymbol{\nu}(t, \cdot)\right|_{\mathscr{BV}}$ denotes the sum of the semi-norm of all the elements in $\boldsymbol{\nu}(t, \cdot)$ [1].

We will use the reformulation in (3.18) to compare the above optical flow based model from [13] with our proposed model. The use of the brightness constancy equation points to using the geometric deformation in (2.4) as a group action in (3.17), i.e., we assume (3.20) holds. It is easy to see that the constraints in (3.18) and (3.24) are equivalent. Hence, the optical flow constraint given by the brightness constancy equation in [13] is equivalent to using diffeomorphisms generated by LDDMM that act through geometric deformation.

By comparison, the primary distinction between (3.18) and (3.24) relates to the selection of the regularization term w.r.t. vector field $\boldsymbol{\nu}(t, \cdot)$. In the model (3.24), one uses the TV functional, so the space of vector fields is assumed to be in $\mathscr{BV}(\Omega)$, which allows for a vector field that is a piecewise-constant vector-valued function distributed on Ω. In contrast, in the model (3.18), the space of vector fields reside in an admissible Hilbert space. Hence, the vector field is a sufficiently smooth vector-valued function distributed on Ω. This guarantees an elastic diffeomorphic deformation, which to some extent mimics the underlying physical mechanisms of organ motion [30, 14].

In addition to the above, both approaches also differ in the selection of regularization term $\mathcal{R}_1$. In (3.18) one only poses restriction on the initial image $f(0, \cdot)$, whereas in (3.24) the whole time trajectory $t \mapsto f(t, \cdot)$ is regularized. The treatment in (3.18) is reasonable to some extent, since regularity properties are assumed to preserved. More precisely, $f(t, \cdot)$ is contained in the same space for all t. For example, if $f(0, \cdot) \in \mathscr{BV}(\Omega)$, then there exists a unique weak solution $f(t, \cdot) = f(0, \cdot) \circ (\phi_{0,t}^{\boldsymbol{\nu}})^{-1}$ to the PDE constraint of (3.18) is $\mathscr{L}^\infty$ in time and in $\mathscr{BV}(\Omega)$ at any t [21, Theorem 4]. Hence, (3.18) has a simpler structure which is also beneficial in implementation.

3.3.2. Diffeomorphic motion models. A diffeomorphic motion model for 4D CT image reconstruction was proposed in [33]. It is based on the LDDMM growth model [31] with a time-continuous model that reads as

$$\min_{\substack{I \in \mathscr{X} \\ \boldsymbol{\nu} \in \mathscr{L}^2_{\mathscr{V}}(\Omega)}} \int_0^1 \left[\mathcal{D}\big(\mathcal{T}_t(\phi_{0,t}^{\boldsymbol{\nu}} \, . \, I), g(t, \cdot)\big) + \mu_2 \big\|\boldsymbol{\nu}(t, \cdot)\big\|_{\mathscr{V}}^2 \right] \mathrm{d}t \tag{3.25}$$
$$\text{s.t. } \phi_{0,t}^{\boldsymbol{\nu}} \text{ solves (3.1).}$$

Compared to (3.11), the above has no regularization term $\mathcal{R}_1$. Another difference relates to the selection on the shape regularization $\mathcal{R}_2$. In (3.25) one uniformly weights the $\big\|\boldsymbol{\nu}(t, \cdot)\big\|_{\mathscr{V}}^2$ term. In contrast, (3.11) uses a non-uniformly weighting (see (3.12) for more clarity) that puts more weight on the previous time instance.

Remark 3.6. Note that in (3.11), we regularize the velocity field more at the beginning, and which is relevant because the template is selected at the initial time $t = 0$, then the

velocity field at time t_i has influence on the geodesic trajectory at $t \geq t_i$. Equivalently, the geodesic trajectory at time t_i depends on the velocity field at $t \leq t_i$. The earlier on we are at the velocity field, the more influence one has on the entire geodesic trajectory. Since the optimal velocity field (3.14) is averagely distributed w.r.t. time t (subsection 3.2.1), the motion of the object is close to a uniform speed. This can also result in a velocity field that is non-vanishing at the end time points (see (3.15) and (3.16)).

Next, by Theorem A.4 it is easy to see that the $\mathscr{L}^2_{\mathscr{V}}(\Omega)$-norm minimizer of (3.25) satisfies

$$\boldsymbol{\nu}(t,\cdot) = \frac{1}{2\mu_2}\int_t^1 \mathcal{K}\Big(\nabla(\phi^{\boldsymbol{\nu}}_{0,t}\,.\,I)\big|D(\phi^{\boldsymbol{\nu}}_{t,\tau})\big|\,\nabla\mathcal{D}_{g_\tau}(\phi^{\boldsymbol{\nu}}_{0,\tau}\,.\,I)(\phi^{\boldsymbol{\nu}}_{t,\tau})\Big)\,\mathrm{d}\tau \quad \text{for } 0 \leq t \leq 1. \tag{3.26}$$

It is not difficult to see from (3.26) that the optimal velocity field is *not averagely distributed w.r.t. time* t. In addition, the minimizer w.r.t. variations of the template satisfies

$$\int_0^1 \big|D(\phi^{\boldsymbol{\nu}}_{0,t})\big|\,\nabla\mathcal{D}_{g_t}(\phi^{\boldsymbol{\nu}}_{0,t}\,.\,I)(\phi^{\boldsymbol{\nu}}_{0,t})\,\mathrm{d}t = 0. \tag{3.27}$$

Combining (3.26) and (3.27) yields $\boldsymbol{\nu}(0,\cdot) = \boldsymbol{\nu}(1,\cdot) = 0$, i.e., the optimal velocity field according to the model (3.25) *vanishes at both the initial and end time points.*

To summarise, both (3.11) and the approach taken in [33] address joint reconstruction and motion estimation. Both approaches model the latter as diffeomorphic deformations generated by velocity fields within the LDDMM framework. A difference is that the regularization of the velocity field in [33] is equally weighted over the entire time trajectory. Next, the optimal velocity field vanishes at both the initial and end time points. This is not the case with the model in (3.11).

A variant to (3.11) based on (2.6) would regularise the template I (image at $t = 0$) instead of its entire time evolution $\phi^{\boldsymbol{\nu}}_{0,t}\,.\,I$. This leads to

$$\begin{aligned}&\min_{\substack{I\in\mathscr{X}\\ \boldsymbol{\nu}\in\mathscr{L}^2_{\mathscr{V}}(\Omega)}} \int_0^1 \Big[\mathcal{D}\big(\mathcal{T}_t(\phi^{\boldsymbol{\nu}}_{0,t}\,.\,I), g(t,\,\cdot\,)\big) + \mu_1\mathcal{R}_1(\phi^{\boldsymbol{\nu}}_{0,t}\,.\,I) + \mu_2\int_0^t \big\|\boldsymbol{\nu}(\tau,\cdot)\big\|^2_{\mathscr{V}}\,\mathrm{d}\tau\Big]\,\mathrm{d}t\\ &\text{s.t. } \phi^{\boldsymbol{\nu}}_{0,t} \text{ solves (3.1).}\end{aligned} \tag{3.28}$$

By Theorem A.4, the optimal velocity field for $0 \leq t < 1$ satisfies

$$\boldsymbol{\nu}(t,\,\cdot\,) = \frac{1}{2\mu_2(1-t)}\int_t^1 \mathcal{K}\Big(\nabla(\phi^{\boldsymbol{\nu}}_{0,t}\,.\,I)\big|D(\phi^{\boldsymbol{\nu}}_{t,\tau})\big|\,\nabla\mathcal{S}_{g_\tau}(\phi^{\boldsymbol{\nu}}_{0,\tau}\,.\,I)(\phi^{\boldsymbol{\nu}}_{t,\tau})\Big)\,\mathrm{d}\tau \tag{3.29}$$

where $\mathcal{S}_{g_t}(f) := \mathcal{D}\Big(\mathcal{T}_t(f), g(t,\,\cdot\,)\Big) + \mu_1\mathcal{R}_1(f)$ for fixed $g(t,\,\cdot\,) \in \mathscr{Y}$ and $f \in \mathscr{X}$. The corresponding optimal template satisfies

$$\int_0^1 \big|D(\phi^{\boldsymbol{\nu}}_{0,t})\big|\,\nabla\mathcal{S}_{g_t}\big(\phi^{\boldsymbol{\nu}}_{0,t}\,.\,I\big)\big(\phi^{\boldsymbol{\nu}}_{0,t}\big)\,\mathrm{d}t = 0. \tag{3.30}$$

Evidently, the above optimal velocity field is also a time average of the integrand. Next, $\boldsymbol{\nu}(1,\,\cdot\,)$ is well-defined at $t = 1$ and reads as

$$\boldsymbol{\nu}(1,\cdot) = \frac{1}{2\mu_2}\mathcal{K}\big(\nabla(\phi^{\boldsymbol{\nu}}_{0,1}\,.\,I)\,\nabla\mathcal{S}_{g_1}(\phi^{\boldsymbol{\nu}}_{0,1}\,.\,I)\big).$$

However, by (3.29) and (3.30) at $t = 0$ we have $\boldsymbol{\nu}(0, \cdot) = 0$. Hence, the optimal velocity field in (3.28) is *averagely distributed w.r.t. time* t, but *vanishing at the initial time point*. In addition, as stated in subsection 3.3.1, we only need to regularize I rather than $\phi^{\boldsymbol{\nu}}_{0,t} \,.\, I$ to some extent, since I and its time evolution reside in the same space. This proposed model has a simpler structure to implement.

To summarise, the comparative analysis points to several advantages that come with using (3.11) over alternative approaches.

4. Numerical implementation.

The first part of the numerical implementation is to derive a time-discretized formulation of (3.11).

4.1. Time-discretized formulation.

Assume the time sampling of data is uniform, i.e., there is a uniform partition of $[0, 1]$ and data is acquired at time points $\{t_i\}_{i=0}^N$ with $t_i = i/N$ for $0 \leq i \leq 1$. We refer to this as the *gating grid* and the time-discretized formulation of the general spatiotemporal inverse problem in (2.1) is the task to recover $t \mapsto f(t, \cdot)$ from data $g(t_i, \cdot) \in \mathscr{Y}$ where

$$g(t_i, \cdot) = \mathcal{T}_{t_i}\big(f(t_i, \cdot)\big) + g_{\text{noise}}(t_i, \cdot) \quad \text{for } i = 1, \dots, N. \tag{4.1}$$

Then, a time-discretized version of (3.11) reads as

$$\begin{aligned} &\min_{\substack{I \in \mathscr{X} \\ \boldsymbol{\nu} \in \mathscr{L}^2_{\mathscr{V}}(\Omega)}} \left\{ \frac{1}{N} \sum_{i=1}^{N} \left[\mathcal{D}\big(\mathcal{T}_{t_i}(\phi^{\boldsymbol{\nu}}_{0,t_i} \,.\, I), g(t_i, \cdot)\big) + \mu_2 \int_0^{t_i} \big\|\boldsymbol{\nu}(\tau, \cdot)\big\|^2_{\mathscr{V}} \, \mathrm{d}\tau \right] + \mu_1 \mathcal{R}_1(I) \right\} \\ &\quad \text{s.t. } \phi^{\boldsymbol{\nu}}_{0,t} \text{ solves (3.1).} \end{aligned} \tag{4.2}$$

Remark 4.1. The time-discretized version (4.2) can be also written such that the image in the first gate is the template:

$$\begin{aligned} &\min_{\substack{I \in \mathscr{X} \\ \boldsymbol{\nu} \in \mathscr{L}^2_{\mathscr{V}}(\Omega)}} \left\{ \frac{1}{N+1} \sum_{i=0}^{N} \left[\mathcal{D}\big(\mathcal{T}_{t_i}(\phi^{\boldsymbol{\nu}}_{0,t_i} \,.\, I), g(t_i, \cdot)\big) + \mu_2 \int_0^{t_i} \big\|\boldsymbol{\nu}(\tau, \cdot)\big\|^2_{\mathscr{V}} \, \mathrm{d}\tau \right] + \mu_1 \mathcal{R}_1(I) \right\} \\ &\quad \text{s.t. } \phi^{\boldsymbol{\nu}}_{0,t} \text{ solves (3.1).} \end{aligned}$$

Since (4.2) contains highly coupled arguments, it is difficult to jointly solve for the template I and the velocity field $\boldsymbol{\nu}$. An approach that is computationally more feasible is to compute I and $\boldsymbol{\nu}$ through an intertwined iterative scheme. More precisely, if the velocity field $\boldsymbol{\nu}$ is given then the spatiotemporal reconstruction problem (4.2) reduces to a static image reconstruction problem: Conversely, if the template I is fixed, then (4.2) reduces to a sequentially indirect image registration problem where we seek the velocity field $\boldsymbol{\nu}$ from time-series data that are indirect observations of the target. Formalising the above, we try to solve (4.2) through the following intertwined iterative scheme $(I^k, \boldsymbol{\nu}^k) \in \mathscr{X} \times \mathscr{L}^2_{\mathscr{V}}(\Omega)$:

$$\begin{cases} I^k \in \underset{I \in \mathscr{X}}{\arg\min} \; \mathcal{J}_{\boldsymbol{\nu}^{k-1}}(I) \\ \boldsymbol{\nu}^k \in \underset{\boldsymbol{\nu} \in \mathscr{L}^2_{\mathscr{V}}(\Omega)}{\arg\min} \; \mathcal{E}_{I^k}(\boldsymbol{\nu}) \end{cases} \tag{4.3}$$

where the functionals $\mathcal{J}_{\boldsymbol{\nu}}\colon \mathscr{X} \to \mathbb{R}$, for given $\boldsymbol{\nu} \in \mathscr{L}^2_{\mathscr{V}}(\Omega)$, and $\mathcal{E}_I\colon \mathscr{L}^2_{\mathscr{V}}(\Omega) \to \mathbb{R}$, for given $I \in \mathscr{X}$, are defined as

$$\mathcal{J}_{\boldsymbol{\nu}}(I) := \frac{1}{N}\sum_{i=1}^{N} \mathcal{D}\big(\mathcal{T}_{t_i}(\phi_{0,t_i}^{\boldsymbol{\nu}^{k-1}} \,.\, I), g(t_i, \,\cdot\,)\big) + \mu_1 \mathcal{R}_1(I) \tag{4.4}$$

$$\mathcal{E}_I(\boldsymbol{\nu}) := \frac{1}{N}\sum_{i=1}^{N} \Big[\mathcal{D}\big(\mathcal{T}_{t_i}(\phi_{0,t_i}^{\boldsymbol{\nu}} \,.\, I_k), g(t_i, \,\cdot\,)\big) + \mu_2 \int_0^{t_i} \big\|\boldsymbol{\nu}(\tau, \cdot)\big\|^2_{\mathscr{V}} \,\mathrm{d}\tau\Big] \tag{4.5}$$
$$\text{s.t. } \phi_{0,t}^{\boldsymbol{\nu}} \text{ solves (3.1)}.$$

To proceed, we consider the special case when the data discrepancy term is the squared $\mathscr{L}^2$-norm as in (2.7) and the spatial regularization is the TV functional as in (2.10). Then (4.2) reads as

$$\min_{\substack{I\in\mathscr{X}\\ \boldsymbol{\nu}\in\mathscr{L}^2_{\mathscr{V}}(\Omega)}} \left\{ \frac{1}{N}\sum_{i=1}^{N} \Big[\|\mathcal{T}_{t_i}(\phi_{0,t_i}^{\boldsymbol{\nu}} \,.\, I) - g(t_i, \,\cdot\,)\|_2^2 + \mu_2 \int_0^{t_i} \big\|\boldsymbol{\nu}(\tau, \cdot)\big\|^2_{\mathscr{V}} \,\mathrm{d}\tau\Big] + \mu_1 \|\nabla I\|_1 \right\} \tag{4.6}$$
$$\text{s.t. } \phi_{0,t}^{\boldsymbol{\nu}} \text{ solves (3.1)}.$$

Correspondingly, (4.5) become

$$\mathcal{J}_{\boldsymbol{\nu}}(I) = \frac{1}{N}\sum_{i=1}^{N} \Big\|\mathcal{T}_{t_i}\big(\phi_{0,t_i}^{\boldsymbol{\nu}} \,.\, I\big) - g(t_i, \,\cdot\,)\Big\|_2^2 + \mu_1 \|\nabla I\|_1 \tag{4.7}$$

$$\mathcal{E}_I(\boldsymbol{\nu}) = \frac{1}{N}\sum_{i=1}^{N} \left[\Big\|\mathcal{T}_{t_i}\big(\phi_{0,t_i}^{\boldsymbol{\nu}} \,.\, I\big) - g(t_i, \,\cdot\,)\Big\|_2^2 + \mu_2 \int_0^{t_i} \big\|\boldsymbol{\nu}(\tau, \cdot)\big\|^2_{\mathscr{V}} \,\mathrm{d}\tau\right] \tag{4.8}$$

Inserting the above into (4.3) yields the following intertwined scheme for solving (4.6):

$$\begin{cases} I^k := \operatorname*{arg\,min}_{I\in\mathscr{X}} \mathcal{J}_{\boldsymbol{\nu}^{k-1}}(I) \quad \text{with } \mathcal{J}_{\boldsymbol{\nu}^{k-1}} \text{ given by (4.7)} \\ \boldsymbol{\nu}^k := \operatorname*{arg\,min}_{\boldsymbol{\nu}\in\mathscr{L}^2_{\mathscr{V}}(\Omega)} \mathcal{E}_{I^k}(\boldsymbol{\nu}) \quad \text{with } \mathcal{E}_{I^k} \text{ given by (4.8)}. \end{cases} \tag{4.9}$$

4.2. Template reconstruction. We henceforth consider geometric deformations where the deformation operator is given as $\phi_{0,t} \,.\, I := I \circ \phi_{0,t}^{-1} = I \circ \phi_{t,0}$. This is a common choice for image registration [7, 9, 36, 13].

Next, consider the problem of computing a minimiser to $\mathcal{J}_{\boldsymbol{\nu}^{k-1}}$ (static reconstruction) in (4.7). This is a non-smooth TV-ℓ_2 minimization problem that is convex when the forward operator $\mathcal{T}_{t_i}\colon \mathscr{X} \to \mathscr{Y}$ is linear. Solving it by a first order method, like a gradient descent scheme as outlined in Algorithm 4.1, requires smoothing the TV-component, i.e., we seek to solve

$$\min_{I\in\mathscr{X}} \left\{ \frac{1}{N}\sum_{i=1}^{N} \big\|\mathcal{T}_{t_i}\big(I \circ \phi_{t_i,0}^{\boldsymbol{\nu}}\big) - g(t_i, \,\cdot\,)\big\|_2^2 + \mu_1 \int_\Omega |\nabla I(x)|_{2,\epsilon} \,\mathrm{d}x \right\}, \tag{4.10}$$

where $|\nabla I(x)|_{2,\epsilon} := \sqrt{\sum_i \big(\partial_i I(x)\big)^2 + \epsilon}$ with $\epsilon > 0$ small, e.g., $\epsilon = 10^{-12}$. This is a frequently used modification of the TV functional in imaging applications [50, 18, 16]. A gradient descent scheme for (4.10) assuming a linear $\mathcal{T}_{t_i}$ reads as

$$I^{k+1} = I^k - \alpha^k \Bigg(\frac{2}{N} \sum_{i=1}^{N} \big|D(\phi^{\boldsymbol{\nu}}_{0,t_i})\big| \mathcal{T}^*_{t_i} \Big(\mathcal{T}_{t_i}(I^k \circ \phi^{\boldsymbol{\nu}}_{t_i,0}) - g(t_i, \cdot) \Big) (\phi^{\boldsymbol{\nu}}_{0,t_i}) + \mu_1 \nabla^* \Big(\frac{\nabla I^k}{|\nabla I^k|_{2,\epsilon}} \Big) \Bigg). \tag{4.11}$$

Here, α^k is the stepsize for the k-th iteration and $\mathcal{T}^*_{t_i}$ is the adjoint of $\mathcal{T}_{t_i}$. The iterative scheme (4.11) is used for updating the template in the intertwined scheme (4.3) for solving (4.6).

Remark 4.2. There are several optimization techniques for solving convex non-smooth problems, like minimising $I \mapsto \mathcal{J}_{\boldsymbol{\nu}}(I)$ in (4.7). Proximal gradient methods [49, 19, 42, 20, 4] are an important class of methods that are suitable for solving non-differentiable convex optimization problems, so they can be applied for solving (4.7). However, evaluating the proximal operator of a function itself involves solving a small convex optimization problem. For this reason, these more advanced methods introduce further auxiliary variables/parameters and they come with slower convergence rates.

4.2.1. Computing diffeomorphic deformations. Updating the template requires access to diffeomorphic deformations $\phi^{\boldsymbol{\nu}}_{t_i,0}$ and $\phi^{\boldsymbol{\nu}}_{0,t_i}$ for $1 \le i \le N$. By definition, $\phi^{\boldsymbol{\nu}}_{s,t}$ solves the flow equation

$$\begin{cases} \partial_t \varphi(t,x) = \boldsymbol{\nu}\big(t, \varphi(t,x)\big) \\ \varphi(s,x) = x \end{cases} \qquad \text{for } x \in \Omega \text{ and } 0 \le s,t \le 1, \tag{4.12}$$

where s is a fixed time point. Integrating w.r.t. time t in (4.12) yields

$$\phi^{\boldsymbol{\nu}}_{s,t} = \mathrm{Id} + \int_s^t \boldsymbol{\nu}\big(\tau, \phi^{\boldsymbol{\nu}}_{s,\tau}\big)\, \mathrm{d}\tau \quad \text{for } 0 \le t \le 1. \tag{4.13}$$

The time interval $[0,1]$ is subdivided uniformly into MN parts thereby forming a discretized time grid that is given as $\tau_j = j/(MN)$ for $j = 0, 1, \ldots, MN$. Evidently, $\tau_{iM} = t_i$ for $i = 0, 1, \ldots, N$, so each subinterval $[t_i, t_{i+1}]$ is segmented into M even parts. The M is named as a factor of discretized time degree. If $M = 1$, then $\tau_i = t_i$, implying that the discretized time grid is consistent with the gating grid. Note also that the different subintervals of the gating grid can be discretized adaptively according to the degree of motions.

Within a short-time interval one can approximate the diffeomorphic deformation with linearized deformations [43]. More precisely, let $s = \tau_j$, $t = \tau_{j-1}$ and τ_{j+1} in (4.13). Then, the small deformations $\phi^{\boldsymbol{\nu}}_{\tau_i,\tau_{i-1}}$ and $\phi^{\boldsymbol{\nu}}_{\tau_i,\tau_{i+1}}$ can be approximated by

$$\phi^{\boldsymbol{\nu}}_{\tau_j,\tau_{j-1}} \approx \mathrm{Id} - \frac{1}{MN} \boldsymbol{\nu}(\tau_j, \cdot), \quad \text{and} \quad \phi^{\boldsymbol{\nu}}_{\tau_j,\tau_{j+1}} \approx \mathrm{Id} + \frac{1}{MN} \boldsymbol{\nu}(\tau_j, \cdot). \tag{4.14}$$

Moreover, by (3.4) we get $\phi^{\boldsymbol{\nu}}_{\tau_j,0} = \phi^{\boldsymbol{\nu}}_{\tau_{j-1},0} \circ \phi^{\boldsymbol{\nu}}_{\tau_j,\tau_{j-1}}$, which combined with (4.14) yields

$$\phi^{\boldsymbol{\nu}}_{\tau_j,0} \approx \phi^{\boldsymbol{\nu}}_{\tau_{j-1},0} \circ \Big(\mathrm{Id} - \frac{1}{MN}\boldsymbol{\nu}(\tau_j, \cdot)\Big) \quad \text{for } j = 1, 2, \ldots, MN. \tag{4.15}$$

This yields the following estimate for $I \circ \phi^{\boldsymbol{\nu}}_{\tau_j,0}$:

$$I \circ \phi^{\boldsymbol{\nu}}_{\tau_j,0} \approx \big(I \circ \phi^{\boldsymbol{\nu}}_{\tau_{j-1},0}\big) \circ \Big(\mathrm{Id} - \frac{1}{MN}\boldsymbol{\nu}(\tau_j, \cdot)\Big) \quad \text{for } j = 1, 2, \ldots, MN \tag{4.16}$$

and $I \circ \phi^{\boldsymbol{\nu}}_{\tau_0,0} = I$. Similarly, (3.4) also implies $\phi^{\boldsymbol{\nu}}_{\tau_j,t_i} = \phi^{\boldsymbol{\nu}}_{\tau_{j+1},t_i} \circ \phi^{\boldsymbol{\nu}}_{\tau_j,\tau_{j+1}}$ for $i \geq 1$, which combined with (4.14) gives the following approximation:

$$\phi^{\boldsymbol{\nu}}_{\tau_j,t_i} \approx \phi^{\boldsymbol{\nu}}_{\tau_{j+1},t_i} \circ \Big(\mathrm{Id} + \frac{1}{MN}\boldsymbol{\nu}(\tau_j, \cdot)\Big) \quad \text{for } j = iM - 1, iM - 2, \ldots, 0 \tag{4.17}$$

and $\phi^{\boldsymbol{\nu}}_{t_i,t_i} = \mathrm{Id}$. To summarize, the deformation between two images at adjacent points in the discretized time grid is approximately represented by a linearized deformation.

4.2.2. Computing mass-preserving deformations. Expression for the gradient of the data discrepancy term for the mass-preserving group action in (2.5) involves terms of the type

$$\big|D\big(\phi^{\boldsymbol{\nu}}_{0,t_i}\big)\big|\mathcal{T}^*_{t_i}\Big(\mathcal{T}_{t_i}\big(I \circ \phi^{\boldsymbol{\nu}}_{t_i,0}\big) - g(t_i, \cdot)\Big) \circ \phi^{\boldsymbol{\nu}}_{0,t_i} \quad \text{for } i \geq 1.$$

Starting with the Jacobian determinant, by (4.17) we get

$$\big|D(\phi^{\boldsymbol{\nu}}_{\tau_j,t_i})\big| \approx \Big(1 + \frac{1}{MN}\operatorname{div}\boldsymbol{\nu}(\tau_j, \cdot)\Big)\big|D(\phi^{\boldsymbol{\nu}}_{\tau_{j+1},t_i})\big| \circ \Big(\mathrm{Id} + \frac{1}{MN}\boldsymbol{\nu}(\tau_j, \cdot)\Big) \tag{4.18}$$

for $j = iM - 1, iM - 2, \ldots, 0$ and where $\big|D(\phi^{\boldsymbol{\nu}}_{t_i,t_i})\big| = 1$. Next, (4.17) also yields the following approximation:

$$\begin{aligned}\mathcal{T}^*_{t_i}\Big(\mathcal{T}_{t_i}\big(I \circ \phi^{\boldsymbol{\nu}}_{t_i,0}\big) - g(t_i, \cdot)\Big) \circ \phi^{\boldsymbol{\nu}}_{\tau_j,t_i} \\ \approx \mathcal{T}^*_{t_i}\Big(\mathcal{T}_{t_i}\big(I \circ \phi^{\boldsymbol{\nu}}_{t_i,0}\big) - g(t_i, \cdot)\Big) \circ \phi^{\boldsymbol{\nu}}_{\tau_{j+1},t_i} \circ \Big(\mathrm{Id} + \frac{1}{MN}\boldsymbol{\nu}(\tau_j, \cdot)\Big)\end{aligned} \tag{4.19}$$

for $j = iM - 1, iM - 2, \ldots, 0$. For simplicity, let

$$\eta^{I,\boldsymbol{\nu}}_{\tau,t} = \big|D\big(\phi^{\boldsymbol{\nu}}_{\tau,t}\big)\big|\mathcal{T}^*_t\Big(\mathcal{T}_t\big(I \circ \phi^{\boldsymbol{\nu}}_{t,0}\big) - g(t, \cdot)\Big) \circ \phi^{\boldsymbol{\nu}}_{\tau,t}. \tag{4.20}$$

Then multiplying (4.18) by (4.19), and using (4.20), $\eta^{I,\boldsymbol{\nu}}_{0,t_i}$ for $i \geq 1$ is computed by

$$\eta^{I,\boldsymbol{\nu}}_{\tau_j,t_i} \approx \Big(1 + \frac{1}{MN}\operatorname{div}\boldsymbol{\nu}(\tau_j, \cdot)\Big)\eta^{I,\boldsymbol{\nu}}_{\tau_{j+1},t_i} \circ \Big(\mathrm{Id} + \frac{1}{MN}\boldsymbol{\nu}(\tau_j, \cdot)\Big) \tag{4.21}$$

for $j = iM - 1, iM - 2, \ldots, 0$ with $\eta^{I,\boldsymbol{\nu}}_{t_i,t_i} = \mathcal{T}^*_{t_i}\Big(\mathcal{T}_{t_i}\big(I \circ \phi^{\boldsymbol{\nu}}_{t_i,0}\big) - g(t_i, \cdot)\Big)$.

Based on the above derivations, the concrete implementation is given as the gradient descent scheme in Algorithm 4.1.

Algorithm 4.1 Gradient descent scheme for minimizing $I \mapsto \mathcal{J}_{\boldsymbol{\nu}}(I)$ in (4.7)

Initialize:
$k \leftarrow 0$.
$t_i \leftarrow \frac{i}{N}$ for $i = 0, 1, \ldots, N$.
$\tau_j \leftarrow \frac{j}{MN}$ for $j = 0, 1, \ldots, MN$.
Given $\boldsymbol{\nu}$.
$I^k \leftarrow I^0$. Here I^0 is a given initial template.
Spatial regularization parameter $\mu_1 > 0$.
Error tolerance $\epsilon_I > 0$, stepsize $\alpha^k = \alpha > 0$, and iteration number $K_I > 0$.
Loop:
Compute $I^k \circ \phi^{\boldsymbol{\nu}}_{\tau_j,0}$ for $1 \leq j \leq MN$ by

$$I^k \circ \phi^{\boldsymbol{\nu}}_{\tau_j,0} \leftarrow \left(I^k \circ \phi^{\boldsymbol{\nu}}_{\tau_{j-1},0}\right) \circ \left(\mathrm{Id} - \frac{1}{MN}\boldsymbol{\nu}(\tau_j, \cdot)\right)$$

with $I^k \circ \phi^{\boldsymbol{\nu}}_{0,0} = I^k$.
Update $\eta^{I^k,\boldsymbol{\nu}}_{t_i,t_i}$ for $1 \leq i \leq N$ by

$$\eta^{I^k,\boldsymbol{\nu}}_{t_i,t_i} \leftarrow \mathcal{T}^*_{t_i}\left(\mathcal{T}_{t_i}(I^k \circ \phi^{\boldsymbol{\nu}}_{t_i,0}) - g(t_i, \cdot)\right)$$

.
Compute $\eta^{I^k,\boldsymbol{\nu}}_{0,t_i}$ for $1 \leq i \leq N$ by

$$\eta^{I^k,\boldsymbol{\nu}}_{\tau_j,t_i} \leftarrow \left(1 + \frac{1}{MN}\operatorname{div}\boldsymbol{\nu}(\tau_j, \cdot)\right)\eta^{I^k,\boldsymbol{\nu}}_{\tau_{j+1},t_i} \circ \left(\mathrm{Id} + \frac{1}{MN}\boldsymbol{\nu}(\tau_j, \cdot)\right)$$

for $j = iM - 1, iM - 2, \ldots, 0$.
Evaluate I^{k+1} by

$$I^{k+1} \leftarrow I^k - \alpha\left(\frac{2}{N}\sum_{i=1}^{N}\eta^{I^k,\boldsymbol{\nu}}_{0,t_i} + \mu_1\,\nabla^*\left(\frac{\nabla I^k}{|\nabla I^k|_{2,\epsilon}}\right)\right).$$

If $\left|I^{k+1} - I^k\right| > \epsilon_I$ and $k < K_I$, then $k \leftarrow k + 1$, **goto** *Loop*.
Output I^{k+1}.

4.3. Velocity field estimation. The aim here is to provide an algorithm for minimising $\mathcal{E}_I$ in (4.8), which amounts to sequential indirect image registration. We will for this purpose use a gradient descent scheme of the form

$$\boldsymbol{\nu}^{k+1} = \boldsymbol{\nu}^k - \beta^k\,\nabla\mathcal{E}_I(\boldsymbol{\nu}^k). \tag{4.22}$$

Here, β^k is the step-size in the k-th iteration, and $\nabla\mathcal{E}_I(\boldsymbol{\nu}^k) \in \mathscr{L}^2_{\mathscr{V}}(\Omega)$ is calculated by (B.1).

The challenge here lies in the computation of this gradient (subsection 4.3.1) and the final algorithm for the gradient descent scheme in (4.22) is given in Algorithm 4.2.

485 **4.3.1. Computing the gradient $\nabla\mathcal{E}_I$.** We begin by introduce notations

486 (4.23) $$h^{I,\boldsymbol{\nu}}_{\tau,t} := \begin{cases} \eta^{I,\boldsymbol{\nu}}_{\tau,t} & 0 \le \tau \le t \le 1 \\ 0 & t < \tau, \end{cases} \quad \text{and} \quad \boldsymbol{\nu}_{\tau,t} := \begin{cases} \boldsymbol{\nu}(\tau,\,\cdot\,) & 0 \le \tau \le t \le 1, \\ 0 & t < \tau. \end{cases}$$

487 Theorem B.1 gives an expression for $\nabla\mathcal{E}_I$ where the kernel function $\mathsf{K}\colon \Omega \times \Omega \to \mathbb{M}^{n\times n}_+$
488 is evaluated on points that do not move as iteration proceeds. By choosing a translation
489 invariant kernel and points on a regular grid in Ω, we can use FFT-based convolution scheme
490 to efficiently evaluate the velocity field at each iteration. This is computationally more feasible
491 than letting the kernel depend on points that move in time as in the shooting method [39, 57].
492 In what follows, we write out the explicit derivations for computing $\nabla\mathcal{E}_I$. As derived in
493 subsection 4.2.1, $I \circ \phi^{\boldsymbol{\nu}}_{\tau_j,0}$ can be approximated by (4.16). The key step is now to update $h^{I,\boldsymbol{\nu}}_{\tau_j,t_i}$
494 for $\{i : t_i \ge \tau_j\}$ in (B.1). We know, by (4.23),

495 (4.24) $$h^{I,\boldsymbol{\nu}}_{\tau_j,t_i} = \eta^{I,\boldsymbol{\nu}}_{\tau_j,t_i} \quad \text{for } t_i \ge \tau_j.$$

496 By using (4.21) for $1 \le i \le N$, we can compute $h^{I,\boldsymbol{\nu}}_{\tau_j,t_i}$ as

497 (4.25) $$h^{I,\boldsymbol{\nu}}_{\tau_j,t_i} \approx \Big(1 + \frac{1}{MN}\operatorname{div}\boldsymbol{\nu}(\tau_j,\,\cdot\,)\Big) h^{I,\boldsymbol{\nu}}_{\tau_{j+1},t_i} \circ \Big(\mathrm{Id} + \frac{1}{MN}\boldsymbol{\nu}(\tau_j,\,\cdot\,)\Big)$$

498 for $j = iM-1, iM-2, \dots, 0$ and with $h^{I,\boldsymbol{\nu}}_{t_i,t_i} = \mathcal{T}^*_{t_i}\Big(\mathcal{T}_{t_i}(I \circ \phi^{\boldsymbol{\nu}}_{t_i,0}) - g(t_i,\,\cdot\,)\Big)$. Hence, at $t = \tau_j$,
499 we get from (B.1) that

500 (4.26) $$\nabla\mathcal{E}_I(\boldsymbol{\nu})(\tau_j, x) = -\frac{2}{N} \sum_{\{i \ge 1 : t_i \ge \tau_j\}} \Big[\int_\Omega \mathsf{K}(x,y) \nabla\big(I \circ \phi^{\boldsymbol{\nu}}_{\tau_j,0}\big)(y) h^{I,\boldsymbol{\nu}}_{\tau_j,t_i}(y)\,\mathrm{d}y - \mu_2 \boldsymbol{\nu}(\tau_j, x) \Big]$$

501 for $0 \le j \le MN$ and $x \in \Omega$. In particular, for $j = MN$ (i.e., $\tau_j = 1$) we have

502 $$\nabla\mathcal{E}_I(\boldsymbol{\nu})(1, x) = -\frac{2}{N} \Big[\int_\Omega \mathsf{K}(x,y) \nabla\big(I \circ \phi^{\boldsymbol{\nu}}_{1,0}\big)(y) h^{I,\boldsymbol{\nu}}_{1,1}(y)\,\mathrm{d}y - \mu_2 \boldsymbol{\nu}(1, x) \Big].$$

503 *Remark* 4.3. It is easy to verify that the optimal solution of the time-discretized version
504 of the proposed model is consistent with that of the time-continuous one. Here the concept
505 of consistency is that the time-discretized solution converges to the time-continuous solution
506 for increasingly fine time discretization. This is however not the case for the diffeomorphic
507 motion model (3.25) in [33]. As an example, at $\tau_j = 1$, the optimal velocity field of the
508 time-discretized problem in [33] satisfies

509 $$\boldsymbol{\nu}(1, x) = \frac{1}{\mu_2} \int_\Omega \mathsf{K}(x,y) \nabla\big(I \circ \phi^{\boldsymbol{\nu}}_{1,0}\big)(y) h^{I,\boldsymbol{\nu}}_{1,1}(y)\,\mathrm{d}y.$$

510 On the other hand, as derived in subsection 3.3.2, the optimal velocity field at $t = 1$ of its
511 time-continuous problem satisfies $\boldsymbol{\nu}(1, x) = 0$. This obviously causes inconsistencies and our
512 consistent approach is an advantage compared to the approach in [33].

513 Finally, Algorithm 4.2 outlines the procedure for computing the gradient descent scheme
514 (4.22) that makes use of the above derivations.

Algorithm 4.2 Gradient descent scheme for minimizing $\boldsymbol{\nu} \mapsto \mathcal{E}_I(\boldsymbol{\nu})$ in (4.8)

Initialize:
$k \leftarrow 0$.
$t_i \leftarrow \frac{i}{N}$ for $i = 0, 1, \ldots, N$.
$\tau_j \leftarrow \frac{j}{MN}$ for $j = 0, 1, \ldots, MN$.
Fixed I.
$\boldsymbol{\nu}^k(\tau_i) \leftarrow \boldsymbol{\nu}^0(\tau_i)$, where $\boldsymbol{\nu}^0$ is a given initial velocity field.
Fixed kernel function $\mathsf{K}(\,\cdot\,,\,\cdot\,)$.
Shape regularization parameter $\mu_2 > 0$.
Error tolerance $\epsilon_{\boldsymbol{\nu}} > 0$, stepsize $\beta^k = \beta > 0$, and maximum iterations $K_{\boldsymbol{\nu}} > 0$.
Loop:
Update $I \circ \phi^{\boldsymbol{\nu}^k}_{\tau_j,0}$ for $1 \le j \le MN$ by

$$I \circ \phi^{\boldsymbol{\nu}^k}_{\tau_j,0} \leftarrow \big(I \circ \phi^{\boldsymbol{\nu}^k}_{\tau_{j-1},0}\big) \circ \Big(\mathrm{Id} - \frac{1}{N}\boldsymbol{\nu}^k(\tau_j,\,\cdot\,)\Big)$$

with $I \circ \phi^{\boldsymbol{\nu}^k}_{0,0} = I$.
Update $h^{I,\boldsymbol{\nu}^k}_{t_i,t_i}$ for $1 \le i \le N$ by

$$h^{I,\boldsymbol{\nu}^k}_{t_i,t_i} \leftarrow \mathcal{T}^*_{t_i}\big(\mathcal{T}_{t_i}(I \circ \phi^{\boldsymbol{\nu}^k}_{t_i,0}) - g(t_i,\,\cdot\,)\big).$$

Compute $h^{I,\boldsymbol{\nu}^k}_{\tau_j,t_i}$ for $1 \le i \le N$ by

$$h^{I,\boldsymbol{\nu}^k}_{\tau_j,t_i} \leftarrow \Big(1 + \frac{1}{MN}\operatorname{div}\boldsymbol{\nu}(\tau_j,\,\cdot\,)\Big) h^{I,\boldsymbol{\nu}^k}_{\tau_{j+1},t_i} \circ \Big(\mathrm{Id} + \frac{1}{MN}\boldsymbol{\nu}^k(\tau_j,\,\cdot\,)\Big)$$

for $j = iM - 1, iM - 2, \ldots, 0$.
Compute $\nabla_{\boldsymbol{\nu}}\,\mathcal{E}_I(\boldsymbol{\nu}^k)(\tau_j,\,\cdot\,)$ (using FFT for convolutions) by

$$\nabla\mathcal{E}_I(\boldsymbol{\nu}^k)(\tau_j, x) \leftarrow -\frac{2}{N}\sum_{\{i\ge 1 : t_i \ge \tau_j\}}\left[\int_\Omega \mathsf{K}(x,y)\,\nabla\big(I \circ \phi^{\boldsymbol{\nu}^k}_{\tau_j,0}\big)(y) h^{I,\boldsymbol{\nu}^k}_{\tau_j,t_i}(y)\,\mathrm{d}y - \mu_2\boldsymbol{\nu}^k(\tau_j,x)\right]$$

for $0 \le j \le MN$.
Update $\boldsymbol{\nu}^k(\tau_j,\,\cdot\,)$ for $0 \le j \le MN$ by:

$$\boldsymbol{\nu}^{k+1}(\tau_j,\,\cdot\,) \leftarrow \boldsymbol{\nu}^k(\tau_j,\,\cdot\,) - \beta\,\nabla\mathcal{E}_I(\boldsymbol{\nu}^k)(\tau_j,\,\cdot\,).$$

If $\big|\boldsymbol{\nu}^{k+1} - \boldsymbol{\nu}^k\big| > \epsilon_{\boldsymbol{\nu}}$ and $k < K_{\boldsymbol{\nu}}$, then $k \leftarrow k+1$, **goto** *Loop*.
Output $\boldsymbol{\nu}^{k+1}$.

4.4. Alternating template reconstruction and velocity field estimation. As described in the beginning of section 4, we aim to solve (4.6) by an iterative scheme where iterates for template image I and velocity field $\boldsymbol{\nu}$ are updated in an alternating manner as in (4.9). Hence, at each iterative step requires solving two sub-problems, one for updating I given $\boldsymbol{\nu}$ (Algorithm 4.1) and the other for updating $\boldsymbol{\nu}$ given I (Algorithm 4.2).

The algorithms for solving the two sub-problems are iterative, so there are inner iterations for each outer iterative step that update the template and velocity field. In order to have a computationally feasible approach, we limit the number of inner iterations to one. The final algorithm for recovering the template and velocity field is presented in Algorithm 4.3. The proposed model is non-convex so a convergence analysis of the algorithm towards a global optima is currently too difficult to obtain.

Algorithm 4.3 Alternately minimizing model (4.8)

1: *Initialize*:
2: Given M, N.
3: $k \leftarrow 0$.
4: $t_i \leftarrow \frac{i}{N}$ for $i = 0, 1, \ldots, N$. This subdivides the time $[0, 1]$ uniformly into N parts.
5: $\tau_j \leftarrow \frac{j}{MN}$ for $j = 0, 1, \ldots, MN$. This subdivides the time interval $[0, 1]$ uniformly into MN parts.
6: Fixed kernel function $\mathsf{K}(\,\cdot\,,\,\cdot\,)$.
7: Given regularization parameters $\mu_1, \mu_2 > 0$.
8: $I^k \leftarrow I^0$, where the template is initialized.
9: $\boldsymbol{\nu}^k(\tau_i) \leftarrow 0$, where the velocity field is initialized to a zero velocity field.
10: Error tolerances $\epsilon_I, \epsilon_{\boldsymbol{\nu}} > 0$, stepsizes $\alpha^k = \alpha > 0, \beta^k = \beta > 0$, and maximum iteration number $K > 0$.
11: *Loop*:
12: Let $\boldsymbol{\nu} = \boldsymbol{\nu}^k$. Perform lines 10–13 in Algorithm 4.1. Output I^{k+1}.
13: Let $I = I^{k+1}$. Perform lines 11–15 in Algorithm 4.2. Output $\boldsymbol{\nu}^{k+1}$.
14: **If** $\left|\boldsymbol{\nu}^{k+1} - \boldsymbol{\nu}^k\right| > \epsilon_{\boldsymbol{\nu}}$ or $\left|I^{k+1} - I^k\right| > \epsilon_I$, and $k < K$, then $k \leftarrow k + 1$, **goto** *Loop*.
15: **Output** I^{k+1}, $\boldsymbol{\nu}^{k+1}$.

Complexity analysis. Here we consider the complexity, including computational cost and memory footprint, of Algorithm 4.3. The main computationally demanding steps at each iteration is located on lines 12–13 (actually lines 10–13 in Algorithm 4.1 and lines 11–15 in Algorithm 4.2), so we restrict our complexity analysis to these parts. For ease of description, we assume $\Omega \subset \mathbb{R}^2$ and the size of the image to be reconstructed is $n \times n$ pixels.

On line 10 of Algorithm 4.1 and line 11 of Algorithm 4.2, we need to update $I \circ \phi^{\boldsymbol{\nu}}_{\tau_j,0}$ for $j = 1, \ldots, MN$. Moreover, each of them should be used to compute the gradient of the objective functional on line 14 of Algorithm 4.2, so they need to be stored at hand. Hence, in these two steps, the computational cost is $O(n^2MN)$ and the memory footprint is $O(n^2MN)$.

For line 11 of Algorithm 4.1 and line 12 of Algorithm 4.2, the $\eta^{I,\boldsymbol{\nu}}_{t_i,t_i}$ (i.e., $h^{I,\boldsymbol{\nu}}_{t_i,t_i}$) need to be updated and then stored for $i = 1, \ldots, N$. The computational cost is $O(n^2N_dN)$, where N_d

is the number of data points. Actually, the N_d is at least proportional to the size of n, which is often $\sqrt{2}nN_v$ with N_v denoting the number of views. Hence, the computational cost scales as $O(n^3 N N_v)$. Since the calculation for the forward and backward projections is on the fly, the required space is not too demanding.

Furthermore, on line 12 of Algorithm 4.1 and line 13 of Algorithm 4.2, for $i = 1, \dots, N$, the $\eta^{I,\boldsymbol{\nu}}_{\tau_j,t_i}$ (i.e., $h^{I,\boldsymbol{\nu}}_{\tau_j,t_i}$) need to be updated and stored for j from $iM - 1$ to 0, then are used to compute the gradient of the objective functional for each time point on line 14 of Algorithm 4.2. Therefore, the computational cost is $O(n^2 M N^2)$. For lines 11-12 of Algorithm 4.1 and lines 12-13 of Algorithm 4.2, the memory footprint is $O(n^2 M N^2)$.

For line 13 of Algorithm 4.1, we need to update I once, the computational cost is $O(n^2 N)$ and the memory footprint is $O(n^2 N)$. At each time point, the FFT is used to compute the gradient of the objective functional on line 14 of Algorithm 4.2. Hence the computational cost for this line is $O(M N^2 n^2 \log n)$. For line 15 of Algorithm 4.2, we need to update a vector field at each time point. Since a vector field would take twice more memory than a scalar field on 2D domain, we spend twice more computational cost to update that. Even so, the computational cost is $O(n^2 M N)$ and the memory footprint is $O(n^2 M N)$.

In summary, for Algorithm 4.3, the computational cost scales as $O(n^3 N)$ and the memory footprint scales as $O(n^2 M N^2)$.

5. Numerical experiments.

In this section, the proposed method for joint image reconstruction and motion estimation is applied to parallel beam tomography with very sparse or highly noisy data in spatiotemporal (e.g., 2D + time) imaging. We use the intensity-preserving group action to consider the involved deformations. The algorithms were implemented in Python using the Operator Discretization Library (http://github.com/odlgroup/odl), which offers GPU parallellized routines for evaluating the ray transform and it adjoint. The source code is available from https://github.com/chongchenmath/SpatiotemporalImaging. The numerical experiments ran on ThinkStation with Xeon E5-2620 v4 2.10 GHz CPU, 64Gb RAM, TITAN Xp GPU, and Ubuntu 16.04 OS. Although not a complete evaluation, the experiments illustrate the performance of the proposed method.

5.1. Spatiotemporal 2D CT.

Image reconstruction in classical static 2D CT amounts to inverting the 2D ray transform. In our spatiotemporal setting, we assume that the temporal variation comes from the object itself that undergoes motion. Hence, the forward operator $\mathcal{T}_t \colon \mathscr{X} \to \mathscr{Y}$, which is the 2D ray transform, is not dependent on time t:

$$\mathcal{T}(f)(\omega, x) := \int_{\mathbb{R}} f(x + s\omega)\, \mathrm{d}s \quad \text{for } \omega \in S^1 \text{ and } x \in \omega^{\perp}.$$

Here, S^1 is the unit circle and (ω, x) encodes a line on $\mathbb{R}^2$ with direction ω through x.

Moreover, consider $\mathscr{V}$ as the space of vector fields that is a RKHS with a reproducing kernel represented by symmetric and positive definite Gaussian function $\mathsf{K} \colon \Omega \times \Omega \to \mathbb{M}^{2\times 2}_{+}$ given as

$$\mathsf{K}(x, y) := \exp\Big(-\frac{1}{2\sigma^2}\|x - y\|_2^2\Big) \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \quad \text{for } x, y \in \mathbb{R}^2 \text{ and } \sigma > 0 \text{ (kernel width)}. \tag{5.1}$$

The images of all gates are supported on Ω. At each gate, the noise-free data per view is

measured by evaluating the 2D parallel beam scanning geometry. Then the additive Gaussian white noise at varying levels is added onto the noise-free data, which leads to the noise data. As in [17], the noise level in data is quantified in terms of signal-to-noise ratio (SNR) defined in logarithmic decibel (dB).

5.2. Test suites and results. The test suites seek to assess the performance against different noise levels, and the sensitivity against various selections of regularization parameters μ_1, μ_2, and kernel width σ. We also compare the proposed method to TV-based static reconstruction method.

5.2.1. Test suite 1: Overview performance. Here we consider a test for evaluating the overview performance. This test uses a multi-object phantom with five gates (i.e., $N = 5$). The used phantom is shown in the last row of Figure 2, which is taken from [17].

The image at each gate is consisting of six separately star-like objects with grey-values over $[0, 1]$, which is digitized using 438×438 pixels. The images of all gates are supported on a fixed rectangular domain $\Omega = [-16, 16] \times [-16, 16]$. At each gate, the noise-free data per view is measured by the 2D parallel beam scanning geometry with even 620 bins, which is supported on the range of $[-24, 24]$. For gate i $(1 \leq i \leq N)$, the scanning views are distributed on $[(i-1)\pi/36, \pi + (i-1)\pi/36]$ uniformly, and the view number is 12. Then three different levels of additive Gaussian white noise are added onto the noise-free data. The resulting SNR are about 4.71dB, 7.7dB, and 14.67dB, respectively. To make it more clear, we show the noise-free and noise projection data at the first view for Gate 1 with respective noise levels in Figure 1.

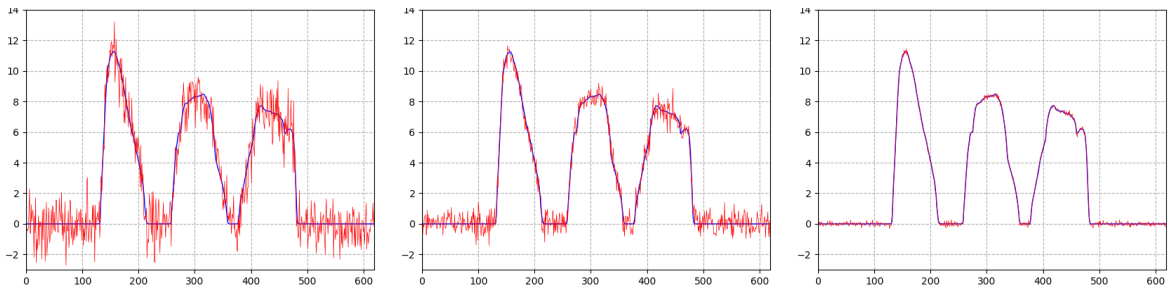

Figure 1: Data at the first view for Gate 1. The left, middle, and right figures show data at the first view for different noise levels 4.71dB, 7.7dB, and 14.67dB, respectively. The blue smooth curve is noise-free data, and the red jagged curve is noisy data.

The factor of discretized time degree is $M = 2$, which is defined in subsection 4.2.1. The kernel width is selected to $\sigma = 2$. The gradient stepsizes are set as $\alpha = 0.01$ and $\beta = 0.05$ respectively, which should not be chosen too large or too small, otherwise could result in the algorithm not convergent or convergent too slow. First we apply Algorithm 4.1 to obtain an initial template image after 50 iterations, then use Algorithm 4.3 to solve the proposed model. Note that the above iteration number is not unchangeable, just needs enough to gain an appropriately initial template for Algorithm 4.3.

The regularization parameters (μ_1, μ_2) are selected as $(0.05, 10^{-7})$ for data noise level

4.71dB, $(0.025, 10^{-7})$ for data noise level 7.7dB, and $(0.01, 10^{-7})$ for data noise level 14.67dB, respectively. The lower SNR, the lager value of μ_1. The maximum iteration number is set to be 200, which should be large enough to guarantee a satisfying result. The runtime for each example is about 29 minutes. The reconstructed results are shown in Figure 2. It is clear that the reconstructed images (rows 1–3) are close to the corresponding ground truth, even though the data SNR is very low.

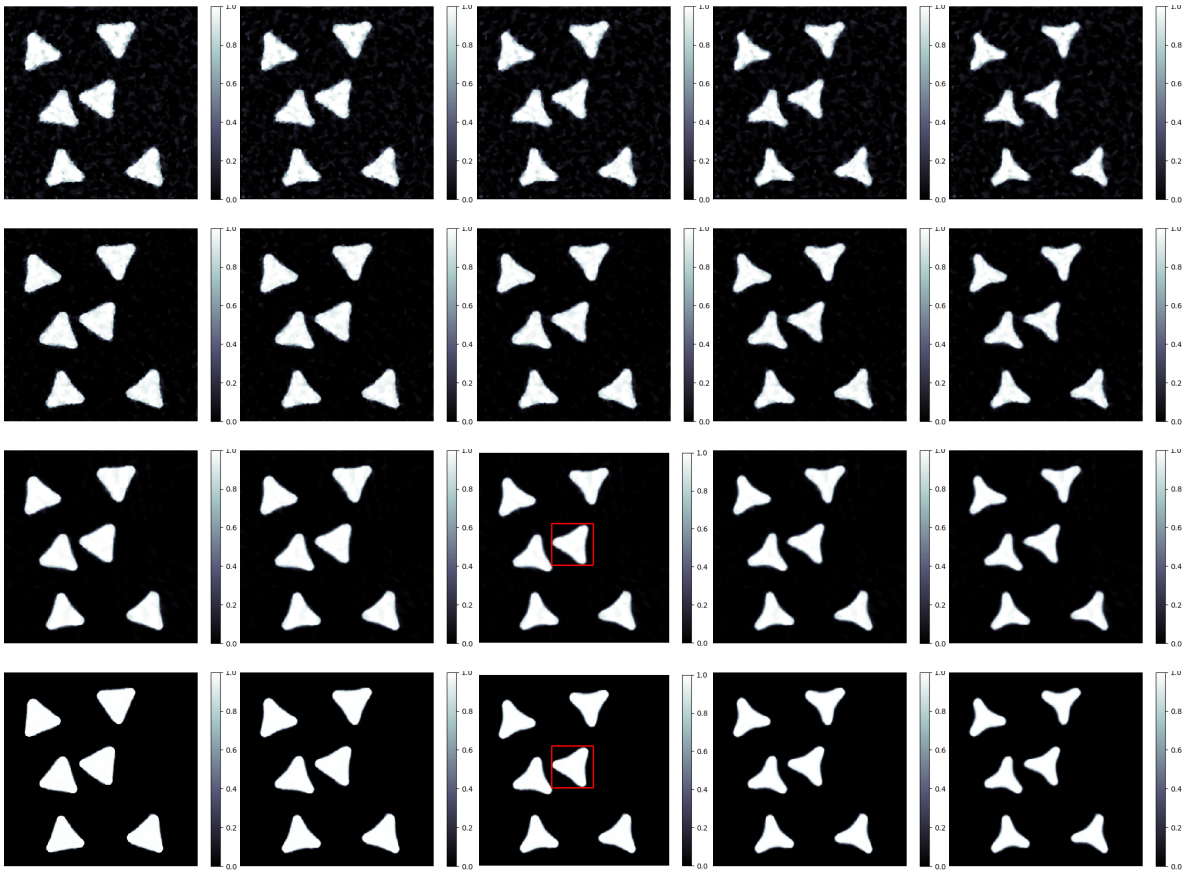

Figure 2: Multi-object phantom. Columns represent the gates and the first three rows are reconstructed spatiotemporal images for the data with noise levels 4.71dB, 7.7dB, and 14.67dB, respectively. The last row shows the ground truth for each gate. The regions of interest (the small boxes in images at the 3rd column in rows 3 and 4) is enlarged in Figure 4.

Apart from the visual perception, the reconstruction is quantitatively compared using structural similarity (SSIM) and peak signal-to-noise ratio (PSNR), which is frequently used to evaluate image quality [59]. The SSIM and PSNR values are tabulated in Table 1. As listed in the above table, the corresponding SSIM and PSNR values are depended on SNR of the data. The higher SNR, the larger values of SSIM and PSNR.

Figure 2	Gate 1	Gate 2	Gate 3	Gate 4	Gate 5
Row 1	0.4069	0.4208	0.4273	0.4305	0.4337
	22.10	23.02	23.27	23.40	23.64
Row 2	0.5934	0.6086	0.6131	0.6149	0.6156
	25.36	27.22	27.37	27.66	27.86
Row 3	0.8411	0.8523	0.8564	0.8576	0.8587
	28.30	31.49	32.48	32.65	32.76

Table 1: SSIM and PSNR values of reconstructed spatiotemporal images for data with varying noise levels, see Figure 2 for the images. Comparisons are made against ground truth. Each table entry has two values, the top being the SSIM and the bottom being the PSNR. The image is shown in the corresponding row/gate in Figure 2.

Comparison against static TV-regularized reconstruction. It is well-known that tomographic reconstruction by TV-regularization outperforms other methods, such as filtered back projection (FBP), when the gradient of the image is sparse. This is furthermore especially notable when data is under-sampled. In our tests we use a phantom (ground truth image) that has sparse gradient, so comparing against static TV-regularized reconstruction pitches our approach against one of the best static reconstruction methods.

For static TV-regularized reconstruction we disregard any temporal evolution, which is equivalent to simplify the spatiotemporal problem into a static problem. The whole tomographic data set will then have 60 projection views. The regularization parameter for static TV-regularized reconstruction is selected depending on the SNR of data in the same way as for spatiotemporal reconstruction.

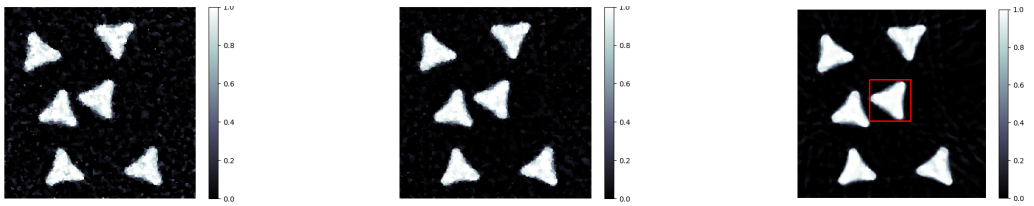

Figure 3: Static TV-regularized reconstructions for the measured data with different noise levels 4.71dB (left), 7.7dB (middle), and 14.67dB (right), respectively. The region of interest in the small box (right) is enlarged in Figure 4.

Reconstructions obtained by static TV-regularized reconstruction are shown in Figure 3, the edges of which become blurring against those by our method. To make it more clear, we enlarge the regions of interest in the small boxes of Figure 2 (column 3, rows 3–4) and Figure 3 for comparison, which are shown in Figure 4.

In addition, the corresponding SSIM and PSNR values are listed in Table 2 and compared to Table 1, the values of SSIM and PSNR for static TV-regularization are lower than those with the proposed method.

Figure 4: Regions of interest for the images in column 3 in Figure 2. Left images shows the region of interest in row 4 (ground truth), middle is for row 3 (the proposed method)) and right is from Figure 3 (static TV-regularization).

Figure 3	Gate 1	Gate 2	Gate 3	Gate 4	Gate 5
Left	0.3012	0.3163	0.3202	0.3146	0.3030
	18.57	19.94	20.42	19.98	18.80
Middle	0.4673	0.4867	0.4910	0.4840	0.4694
	20.44	22.82	23.76	22.90	20.79
Right	0.6004	0.6239	0.6291	0.6212	0.6029
	21.42	24.71	26.40	25.00	21.95

Table 2: SSIM and PSNR values of TV-regularized reconstructions compared to each ground truth. Data is from Gates 1–5 with varying noise levels, see Figure 3 for the images. Each entry has two values, where the upper is the value of SSIM and the bottom is the value of PSNR.

5.2.2. Test suite 2: Sensitivity against selections of regularization parameters. To solve the proposed model, three regularization parameters μ_1, μ_2 and σ need to be selected. Hence the sensitivity test should be concerned against the selections of these parameters.

As shown in the last row of Figure 5, a heart phantom with four gates (i.e., $N = 4$) is used in this test, which is originated from [31]. The image at each gate is consisting of a heart-like object with grey-values in $[0, 1]$, which is digitized using 120×120 pixels. The images of all gates are supported on a fixed rectangular domain $\Omega = [-4.5, 4.5] \times [-4.5, 4.5]$. At each gate, the noise-free data per view is measured by evaluating the 2D parallel beam scanning geometry with uniform 170 bins, which is supported on the range of $[-6.4, 6.4]$. Then the additive Gaussian white noise is added onto the noise-free data. The resulting SNR is about 14.9dB. For gate i $(1 \leq i \leq N)$, the scanning views are distributed on $[(i-1)\pi/5, \pi+(i-1)\pi/5]$ evenly, which totally has 20 views. The factor of discretized time degree is $M = 2$. The gradient stepsizes are set as $\alpha = 0.01$ and $\beta = 0.05$, respectively.

We first employ Algorithm 4.1 to gain an initial template after 50 iterations, then use Algorithm 4.3 to solve the proposed model. With selecting different values for regularization parameters, after 200 iterations, the reconstructed results are obtained, as shown in Figure 5. The runtime for each example is about 3 minutes. The detailed selections of varying parameters can be referred to the caption of Figure 5.

For comparison, we also present the result for static TV-regularized reconstruction in

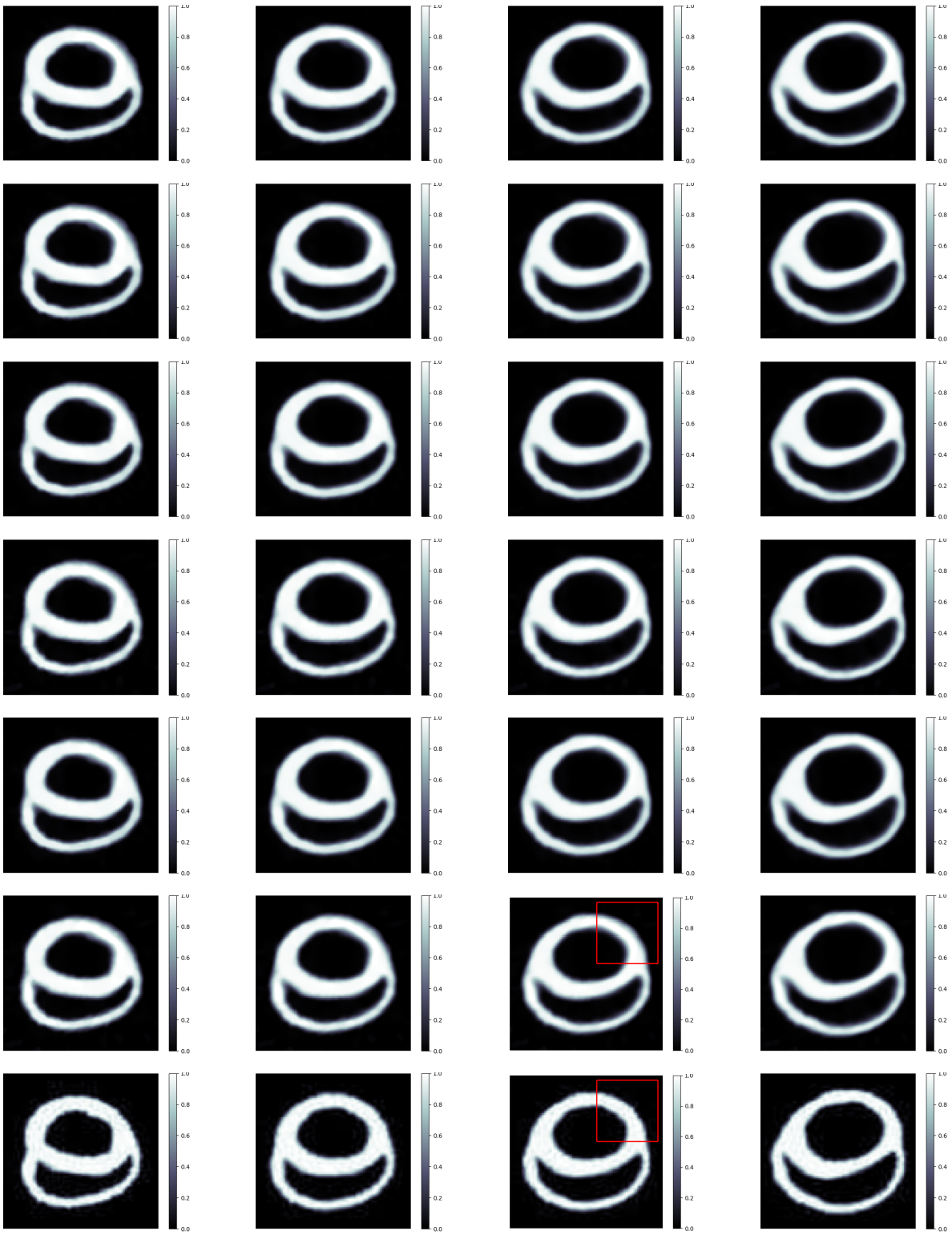

Figure 5: Heart phantom. The columns are the 4 gates and the first 6 rows are reconstructed spatiotemporal images with parameter pairs (μ_1, μ_2, σ) chosen as $(0.01, 10^{-7}, 1.0)$, $(0.01, 10^{-6}, 1.0)$, $(0.01, 10^{-7}, 0.5)$, $(0.005, 10^{-7}, 0.5)$, $(0.01, 10^{-6}, 0.5)$, and $(0.005, 10^{-6}, 0.5)$. The last row shows the ground truth for each gate. The regions of interest in the small boxes (column 3, rows 6–7) are enlarged in Figure 6.

Figure 6 as we did in the first test. As shown in Figure 5, the related reconstructed results are almost the same and close to the counterpart ground truth. However, the reconstructed result by static TV-regularization in Figure 6 is severely degraded. To illustrate this, we enlarge the regions of interest in the small boxes of Figure 5 (column 3, rows 6-7) and Figure 6 for comparison, which are shown in Figure 6.

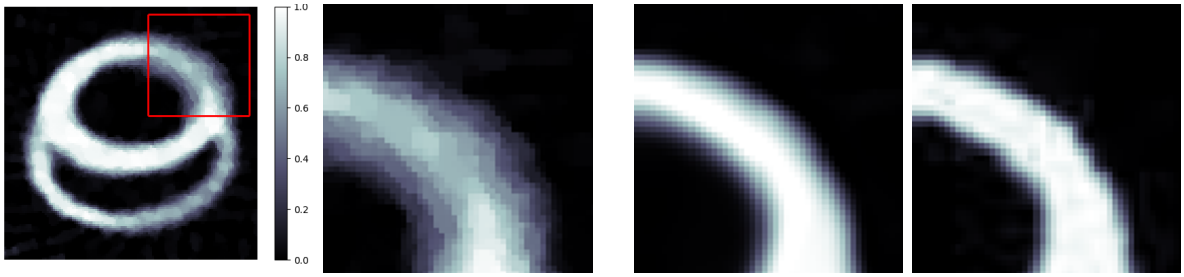

Figure 6: Leftmost image shows reconstruction using static TV-regularization from the measured data with comparable noise level with the region of interest shown separately in the image ot the right. The next image to the right is the same region taken from column 3, row 6 in Figure 5 (the proposed method). Finally, the rightmost image is from column 3, row 7 in Figure 5 (ground truth).

Furthermore, SSIM and PSNR values are reported in Table 3. As can be seen, the SSIM and PSNR values for the proposed method are better than those obtained by static TV-regularized reconstruction. Furthermore, they are quite similar for different parameter choices.

As shown in Table 3, these values are a little bit decreased when the value of kernel parameter σ is changed from 1.0 to 0.5 with fixed μ_1 and μ_2, as compared the values between row 1 and row 3, also row 2 and row 4, for instance. Therefore, this test demonstrates that to some extent the proposed method is not sensitive to the precise selection of the regularization parameters under the visual perception and the quantitative comparison (SSIM and PSNR). However, those values are selected too big or too small, which would causes over- or under-regularized results.

As indicated in subsection 3.3, the optimal velocity field of the proposed model is nonvanishing at both the initial and the end time points. To verify this standpoint in numerical way, we display the computed optimal velocity field at times $t = 0$ and $t = 1$ in Figure 7 for the example with parameter pair $(0.005, 10^{-7}, 0.5)$ on Row 4 of Figure 5.

6. Conclusions and the future work. A general framework of variational model has been investigated for joint image reconstruction and motion estimation in spatiotemporal imaging, which is based on the deformable templates from shape theory. Along this framework, we proposed a new variational model for solving the above joint problem using the principle of LDDMM. The proposed model is equivalent to a PDE-constrained optimal control problem. Based on the equivalency, we made a mathematical comparison against the joint TV-TV optical flow based model [13], which showed that our method can guarantee elastically diffeomorphic deformations, and is of benefit to the practical computation additionally.

	Gate 1	Gate 2	Gate 3	Gate 4
Row 1	0.8928	0.9382	0.9340	0.9235
	24.25	28.44	27.64	26.28
Row 2	0.8960	0.9415	0.9346	0.9234
	24.30	28.47	27.67	26.37
Row 3	0.9103	0.9497	0.9459	0.9343
	25.33	29.41	28.97	27.78
Row 4	0.8940	0.9368	0.9361	0.9291
	25.13	29.06	28.60	27.65
Row 5	0.9087	0.9472	0.9462	0.9336
	25.14	29.30	28.83	27.68
Row 6	0.8884	0.9339	0.9358	0.9295
	25.23	29.06	28.65	27.74
Static TV	0.5641	0.7310	0.7458	0.5969
	14.09	19.09	18.96	14.01

Table 3: Rows 1–6: SSIM and PSNR values of reconstructed spatiotemporal images for varying values of the regularization parameters μ_1, μ_2, and the kernel width σ. Each table entry has two values, the top being the SSIM and the bottom being the PSNR. The image is shown in the corresponding row/gate in Figure 5. The bottom row reports SSIM and PSNR values of TV-regjularized reconstructions compared to each ground truth from gates 1–4 by the measured data with comparable noise level.

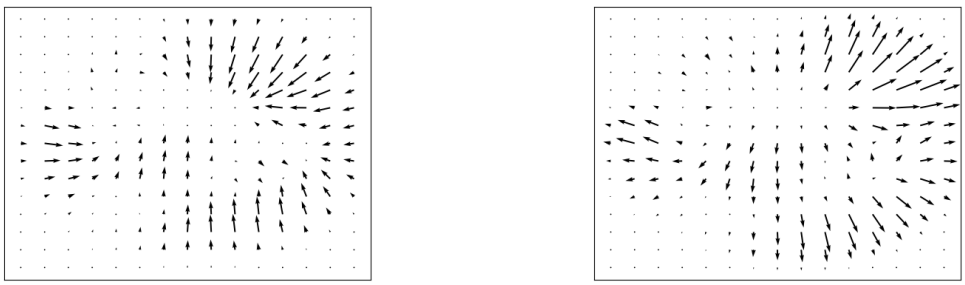

Figure 7: The computed optimal velocity fields at times $t = 0$ (left) and $t = 1$ (right) for the example with parameter pair $(0.005, 10^{-7}, 0.5)$ on row 4 in Figure 5.

Furthermore, the theoretical comparison was also performed between the proposed model and other diffeomorphic motion models, which demonstrated that the optimal velocity field of our model is distributed w.r.t. time t averagely, and non-vanishing at both the initial and the end time points. We also presented an efficient computational method for the time-discretized formulation and showed that its optimal solution is consistent with that of the time-continuous one. This is not the case for the diffeomorphic motion model in [33].

An alternating gradient descent algorithm was designed to solve the time-discretized proposed model, where the main calculations were only based on the easily-implemented linearized deformations. For spatiotemporal (2D space + time) parallel beam tomographic imaging, the computational cost of the algorithm is then $O(n^3 N N_v)$ and its memory requirement scales as $O(n^2 M N^2)$. With Algorithm 4.3, we have evaluated the performance of the proposed model in dealing with the 2D space + time tomography in the case of very sparse and/or highly noisy data. As shown in these visual and quantitative results, the new method can yield reconstructed spatiotemporal images of high quality for the above difficult problems.

The future work will focus on the theoretical analysis of the proposed model, such as the existence and uniqueness of the solution, the convergence analysis of the proposed algorithm, and its extensions and applications to more complicated modalities in spatiotemporal imaging.

Appendix A. Optimality conditions.

To derive optimality conditions for (3.11), we begin with the following result, which is proved in [60].

Lemma A.1. *Let $\boldsymbol{\nu}, \boldsymbol{\eta} \in \mathscr{L}^2_{\mathscr{V}}(\Omega)$, $\phi^{\boldsymbol{\nu}}_{0,t}$ denotes the solution to the ODE in (3.1) with given $\boldsymbol{\nu}$ at time t, and $\phi^{\boldsymbol{\nu}}_{s,t}$ is defined as in (3.4). Then,*

$$\frac{d}{d\epsilon}\phi^{\boldsymbol{\nu}+\epsilon\boldsymbol{\eta}}_{s,t}(x)\Big|_{\epsilon=0} = \int_s^t D\big(\phi^{\boldsymbol{\nu}}_{\tau,t}\big)\big(\phi^{\boldsymbol{\nu}}_{s,\tau}(x)\big)\Big(\boldsymbol{\eta}\big(\tau, \phi^{\boldsymbol{\nu}}_{s,\tau}(x)\big)\Big)\,\mathrm{d}\tau \quad \text{for } x \in \Omega \text{ and } 0 \le s,t \le 1. \tag{A.1}$$

Lemma A.1 can be used to prove the following statement:

Lemma A.2. *Let the assumptions in Lemma A.1 hold and assume $I \in \mathscr{L}^2(\Omega, \mathbb{R})$ is differentiable. Then, under the group action in (2.4) we have*

$$\frac{d}{d\epsilon}\big(\phi^{\boldsymbol{\nu}+\epsilon\boldsymbol{\eta}}_{0,t} \,.\, I\big)(x)\Big|_{\epsilon=0} = -\int_0^t \Big\langle \nabla\big(\phi^{\boldsymbol{\nu}}_{0,\tau} \,.\, I\big)\big(\phi^{\boldsymbol{\nu}}_{t,\tau}(x)\big), \boldsymbol{\eta}\big(\tau, \phi^{\boldsymbol{\nu}}_{t,\tau}(x)\big)\Big\rangle_{\mathbb{R}^n} \mathrm{d}\tau \quad \text{for } x \in \Omega. \tag{A.2}$$

Proof. By the chain rule we get

$$\frac{d}{d\epsilon}\big(\phi^{\boldsymbol{\nu}+\epsilon\boldsymbol{\eta}}_{0,t} \,.\, I\big)(x)\Big|_{\epsilon=0} = \Big\langle \nabla I\big(\phi^{\boldsymbol{\nu}}_{t,0}(x)\big), \frac{d}{d\epsilon}\phi^{\boldsymbol{\nu}+\epsilon\boldsymbol{\eta}}_{t,0}(x)\Big|_{\epsilon=0}\Big\rangle_{\mathbb{R}^n}. \tag{A.3}$$

Using Lemma A.1, we know

$$\frac{d}{d\epsilon}\phi^{\boldsymbol{\nu}+\epsilon\boldsymbol{\eta}}_{t,0}(x)\Big|_{\epsilon=0} = -\int_0^t D\big(\phi^{\boldsymbol{\nu}}_{\tau,0}\big)\big(\phi^{\boldsymbol{\nu}}_{t,\tau}(x)\big)\Big(\boldsymbol{\eta}\big(\tau, \phi^{\boldsymbol{\nu}}_{t,\tau}(x)\big)\Big)\,\mathrm{d}\tau. \tag{A.4}$$

Inserting (A.4) into (A.3), we immediately prove (A.2). ∎

The following result is a direct consequence of the above definition and Lemma A.2.

Lemma A.3. *Let the assumptions in Lemma A.2 hold and define the data discrepancy functional $\mathcal{D}_{g_t} : \mathscr{X} \to \mathbb{R}$ as in (3.13). If $\mathcal{D}_{g_t}$ is differentiable, then*

$$\begin{aligned}\frac{d}{d\epsilon}\mathcal{D}_{g_t}\big(\phi^{\boldsymbol{\nu}+\epsilon\boldsymbol{\eta}}_{0,t} \,.\, I\big)\Big|_{\epsilon=0} \\ = \int_0^t \Big\langle -\big|D(\phi^{\boldsymbol{\nu}}_{\tau,t})\big|\,\nabla\mathcal{D}_{g_t}\big(\phi^{\boldsymbol{\nu}}_{0,t} \,.\, I\big)\big(\phi^{\boldsymbol{\nu}}_{\tau,t}\big)\nabla\big(\phi^{\boldsymbol{\nu}}_{0,\tau} \,.\, I\big), \boldsymbol{\eta}(\tau, \cdot)\Big\rangle_{\mathscr{L}^2(\Omega,\mathbb{R}^n)} \mathrm{d}\tau.\end{aligned} \tag{A.5}$$

We are now ready to characterize optimality conditions for (3.11).

Theorem A.4. *Let the assumptions in Lemma A.3 hold and* $\mathcal{L}\colon \mathscr{X} \times \mathscr{L}^2_{\mathscr{V}}(\Omega) \to \mathbb{R}$ *denotes the objective functional in (3.11), i.e.,*

$$\mathcal{L}(I, \boldsymbol{\nu}) := \int_0^1 \left[\mathcal{D}_{g_t}\big(\phi^{\boldsymbol{\nu}}_{0,t} \,.\, I\big) + \mu_2 \int_0^t \big\|\boldsymbol{\nu}(\tau, \cdot)\big\|^2_{\mathscr{V}} \,\mathrm{d}\tau \right] \mathrm{d}t + \mu_1 \mathcal{R}_1(I). \tag{A.6}$$

Assuming that the regularization term $\mathcal{R}_1$ *is differentiable, and* $\mathscr{V}$ *is a RKHS with a reproducing kernel* $\mathsf{K}\colon \Omega \times \Omega \to \mathbb{M}^{n\times n}_+$. *Then, the optimality conditions for (3.11) read as*

$$\nabla_{\boldsymbol{\nu}} \mathcal{L}(I, \boldsymbol{\nu})(t, \,\cdot\,) = 0 \quad \text{and} \quad \nabla_I \mathcal{L}(I, \boldsymbol{\nu}) = 0. \tag{A.7}$$

Here, the $\mathscr{L}^2_{\mathscr{V}}(\Omega)$*–gradient w.r.t. the velocity field* $\boldsymbol{\nu}$ *of* $\mathcal{L}(I, \,\cdot\,)\colon \mathscr{L}^2_{\mathscr{V}}(\Omega) \to \mathbb{R}$ *is*

$$\nabla_{\boldsymbol{\nu}} \mathcal{L}(I, \boldsymbol{\nu})(t, \,\cdot\,) = \mathcal{K}\left(-\nabla(\phi^{\boldsymbol{\nu}}_{0,t} \,.\, I) \int_t^1 \big|D(\phi^{\boldsymbol{\nu}}_{t,\tau})\big| \, \nabla \mathcal{D}_{g_\tau}\big(\phi^{\boldsymbol{\nu}}_{0,\tau} \,.\, I\big)\big(\phi^{\boldsymbol{\nu}}_{t,\tau}\big) \,\mathrm{d}\tau \right) + 2\mu_2(1-t)\boldsymbol{\nu}(t, \,\cdot\,) \tag{A.8}$$

for $0 \le t \le 1$ *with* $\mathcal{K}(\varphi) := \int_\Omega \mathsf{K}(\,\cdot\,, y)\varphi(y)\,\mathrm{d}y$. *Furthermore, the gradient w.r.t. the template* I *of* $\mathcal{L}(\,\cdot\,, \boldsymbol{\nu})\colon \mathscr{X} \to \mathbb{R}$ *is*

$$\nabla_I \mathcal{L}(I, \boldsymbol{\nu}) = \int_0^1 \big|D(\phi^{\boldsymbol{\nu}}_{0,t})\big| \, \nabla \mathcal{D}_{g_t}\big(\phi^{\boldsymbol{\nu}}_{0,t} \,.\, I\big)\big(\phi^{\boldsymbol{\nu}}_{0,t}\big) \,\mathrm{d}t + \mu_1 \nabla \mathcal{R}_1(I), \tag{A.9}$$

where $\nabla \mathcal{R}_1$ *denotes the gradient of* $\mathcal{R}_1\colon \mathscr{X} \to \mathbb{R}$.

Proof. Applying the result in Lemma A.3, we immediately have

$$\frac{d}{d\epsilon}\mathcal{L}(I, \boldsymbol{\nu} + \epsilon\boldsymbol{\eta})\Big|_{\epsilon=0} = \int_0^1 \int_0^t \Big\langle -\big|D(\phi^{\boldsymbol{\nu}}_{\tau,t})\big| \, \nabla \mathcal{D}_{g_t}\big(\phi^{\boldsymbol{\nu}}_{0,t} \,.\, I\big)\big(\phi^{\boldsymbol{\nu}}_{\tau,t}\big) \nabla(\phi^{\boldsymbol{\nu}}_{0,\tau} \,.\, I), \boldsymbol{\eta}(\tau, \,\cdot\,) \Big\rangle_{\mathscr{L}^2(\Omega, \mathbb{R}^n)} \mathrm{d}\tau \,\mathrm{d}t + 2\mu_2 \int_0^1 \int_0^t \big\langle \boldsymbol{\nu}(\tau, \,\cdot\,), \boldsymbol{\eta}(\tau, \,\cdot\,) \big\rangle_{\mathscr{V}} \,\mathrm{d}\tau \,\mathrm{d}t.$$

Changing the order of integration in the above equation gives

$$\begin{aligned} &\frac{d}{d\epsilon}\mathcal{L}(I, \boldsymbol{\nu} + \epsilon\boldsymbol{\eta})\Big|_{\epsilon=0} \\ &\quad = \int_0^1 \Big\langle -\nabla(\phi^{\boldsymbol{\nu}}_{0,\tau} \,.\, I) \int_\tau^1 \big|D(\phi^{\boldsymbol{\nu}}_{\tau,t})\big| \, \nabla \mathcal{D}_{g_t}\big(\phi^{\boldsymbol{\nu}}_{0,t} \,.\, I\big)\big(\phi^{\boldsymbol{\nu}}_{\tau,t}\big) \,\mathrm{d}t, \boldsymbol{\eta}(\tau, \,\cdot\,) \Big\rangle_{\mathscr{L}^2(\Omega, \mathbb{R}^n)} \mathrm{d}\tau \\ &\qquad + 2\mu_2 \int_0^1 \big\langle (1-\tau)\boldsymbol{\nu}(\tau, \,\cdot\,), \boldsymbol{\eta}(\tau, \,\cdot\,) \big\rangle_{\mathscr{V}} \,\mathrm{d}\tau. \end{aligned} \tag{A.10}$$

As $\mathscr{V}$ is a RKHS with a reproducing kernel represented by $\mathsf{K}\colon \Omega \times \Omega \to \mathbb{M}^{n\times n}_+$, then

$$\langle \boldsymbol{v}, \boldsymbol{u} \rangle_{\mathscr{L}^2(\Omega, \mathbb{R}^n)} = \left\langle \int_\Omega \mathsf{K}(\,\cdot\,, y)\boldsymbol{v}(y) \,\mathrm{d}y, \boldsymbol{u} \right\rangle_{\mathscr{V}} \quad \text{for } \boldsymbol{v}, \boldsymbol{u} \in \mathscr{V}. \tag{A.11}$$

Combining (A.10) with (A.11) proves (A.8). Finally, the results in (A.9) and (A.7) are rather straightforward to obtain, so we omit their proofs. ∎

Appendix B. First-order variation of $\mathcal{E}_I$.

Theorem B.1. *Let the assumptions in Lemma A.2 hold and $\mathcal{E}_I \colon \mathscr{L}^2_{\mathscr{V}}(\Omega) \to \mathbb{R}$ is given as in (4.8). Assume furthermore that $\mathscr{V}$ is a RKHS with a reproducing kernel $\mathsf{K} \colon \Omega \times \Omega \to \mathbb{M}^{n\times n}_+$. Then, the $\mathscr{L}^2_{\mathscr{V}}(\Omega)$–gradient of $\mathcal{E}_I$ is*

$$\nabla \mathcal{E}_I(\boldsymbol{\nu})(t,x) = -\frac{2}{N}\int_\Omega \mathsf{K}(x,y)\nabla(I\circ\phi^{\boldsymbol{\nu}}_{t,0})(y) \sum_{\{i\geq 1 : t_i \geq t\}} h^{I,\boldsymbol{\nu}}_{t,t_i}(y)\,\mathrm{d}y + \frac{2\mu_2}{N}\sum_{\{i\geq 1 : t_i\geq t\}} \boldsymbol{\nu}_{t,t_i}(x), \tag{B.1}$$

for $0 \leq t \leq 1$ and $x \in \Omega$.

Proof. From Lemma A.2 it is not difficult to derive

$$\begin{aligned}
\frac{d}{d\epsilon}\mathcal{E}_I(\boldsymbol{\nu}+\epsilon\boldsymbol{\eta})\Big|_{\epsilon=0} &= \frac{1}{N}\sum_{i=1}^N \int_0^{t_i} \Big\langle -2\eta^{I,\boldsymbol{\nu}}_{\tau,t_i}\nabla(I\circ\phi^{\boldsymbol{\nu}}_{\tau,0}), \boldsymbol{\eta}(\tau,\,\cdot\,)\Big\rangle_{\mathscr{L}^2(\Omega,\mathbb{R}^n)}\mathrm{d}\tau \\
&\quad + \frac{\mu_2}{N}\sum_{i=1}^N \int_0^{t_i}\Big\langle 2\boldsymbol{\nu}(\tau,\,\cdot\,),\boldsymbol{\eta}(\tau,\,\cdot\,)\Big\rangle_{\mathscr{V}}\mathrm{d}\tau \\
&= \int_0^1 \Big\langle -\frac{2}{N}\sum_{i=1}^N h^{I,\boldsymbol{\nu}}_{\tau,t_i}\nabla(I\circ\phi^{\boldsymbol{\nu}}_{\tau,0}),\boldsymbol{\eta}(\tau,\,\cdot\,)\Big\rangle_{\mathscr{L}^2(\Omega,\mathbb{R}^n)}\mathrm{d}\tau \\
&\quad + \int_0^1 \Big\langle \frac{2\mu_2}{N}\sum_{i=1}^N \boldsymbol{\nu}_{\tau,t_i}(\,\cdot\,),\boldsymbol{\eta}(\tau,\,\cdot\,)\Big\rangle_{\mathscr{V}}\mathrm{d}\tau \\
&= \int_0^1 \Big\langle -\frac{2}{N}\sum_{\{i\geq 1:t_i\geq t\}} h^{I,\boldsymbol{\nu}}_{t,t_i}\nabla(I\circ\phi^{\boldsymbol{\nu}}_{t,0}),\boldsymbol{\eta}(t,\,\cdot\,)\Big\rangle_{\mathscr{L}^2(\Omega,\mathbb{R}^n)}\mathrm{d}t \\
&\quad + \int_0^1 \Big\langle \frac{2\mu_2}{N}\sum_{\{i\geq 1:t_i\geq t\}} \boldsymbol{\nu}_{t,t_i}(\,\cdot\,),\boldsymbol{\eta}(t,\,\cdot\,)\Big\rangle_{\mathscr{V}}\mathrm{d}t.
\end{aligned}$$

The last two equations are obtained by inserting (4.23). Combining the above with (A.11) proves (B.1). ∎

Acknowledgments. The authors would like to thank Alain Trouvé for his helpful discussions. The authors also would like to express their gratitude to the anonymous reviewers for their comments, which helped to improve the presentation of this paper.

REFERENCES

[1] G. Aubert and P. Kornprobst, *Mathematical Problems in Image Processing: Partial Differential Equations and the Calculus of Variations*, vol. 147 of Applied Mathematical Sciences, Springer-Verlag, 2002.

[2] W. Bai and M. Brady, *Regularized B-spline deformable registration for respiratory motion correction in PET images*, Phys. Med. Biol., 54 (2009), pp. 2719–2736.

[3] W. Bai and M. Brady, *Motion correction and attenuation correction for respiratory gated PET images*, IEEE Trans. Med. Imaging, 30 (2011), pp. 351–365.

[4] H. H. Bauschke and P. L. Combettes, *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*, CMS Books in Mathematics, Springer, 2ed ed., 2017.

[5] F. M. Beg, M. I. Miller, A. Trouvé, and L. Younes, *Computing large deformation metric mappings via geodesic flow of diffeomorphisms*, International Journal of Computer Vision, 61 (2005), pp. 139—157.

[6] R. Bhagalia, J. Pack, S. Miller, and M. Iatrou, *Nonrigid registration-based coronary artery motion correction for cardiac computed tomography*, Med. Phys., 39 (2012), pp. 4245–4254.

[7] M. Blume, N. Navab, and M. Rafecas, *Joint image and motion reconstruction for PET using a B-spline motion model*, Phys. Med. Biol., 57 (2012), pp. 8249–8270.

[8] M. Brehm, P. Paysan, M. Oelhafen, and M. Kachelriess, *Artifact-resistant motion estimation with a patient-specific artifact model for motion-compensated cone-beam CT*, Med. Phys., 40 (2013), p. 101913.

[9] M. Brehm, P. Paysan, M. Oelhafen, P. Kunz, and M. Kachelriess, *Self-adapting cyclic registration for motion-compensated cone-beam CT in image-guided radiation therapy*, Med. Phys., 39 (2012), pp. 7603–7618.

[10] M. Brehm, S. Sawall, J. Maier, S. Sauppe, and M. Kachelriess, *Cardiorespiratory motion-compensated micro-CT image reconstruction using an artifact model-based motion estimation*, Med. Phys., 42 (2015), pp. 1948–1958.

[11] M. Bruveris and D. D. Holm, *Geometry of image registration: The diffeomorphism group and momentum maps*, in Geometry, Mechanics, and Dynamics: The Legacy of Jerry Marsden, C. D. E., D. D. Holm, G. Patrick, and T. Ratiu, eds., vol. 73 of Fields Institute Communications, Springer-Verlag, 2015, pp. 19–56.

[12] M. Bruveris and F.-X. Vialard, *On completeness of groups of diffeomorphisms*, J. Eur. Math. Soc., 19 (2017), pp. 1507–1544.

[13] M. Burger, H. Dirks, and C. Schönlieb, *A Variational Model for Joint Motion Estimation and Image Reconstruction*, SIAM Journal on Imaging Sciences, 11 (2018), pp. 94–128.

[14] M. Burger, J. Modersitzki, and L. Ruthotto, *A Hyperelastic Regularization Energy for Image Registration*, SIAM J. Sci. Comput., 35 (2013), pp. B132–B148.

[15] F. Büther, M. Dawood, L. Stegger, F. Wübbeling, M. Schäfers, O. Schober, and K. P. Schäfers, *List mode-driven cardiac and respiratory gating in PET*, J. Nucl. Med., 50 (2009), pp. 674–681.

[16] B. Chen, Z. Zhang, E. Y. Sidky, D. Xia, and X. Pan, *Image reconstruction and scan configurations enabled by optimization-based algorithms in multispectral CT*, Phys. Med. Biol., 62 (2017), pp. 8763–8793.

[17] C. Chen and O. Öktem, *Indirect Image Registration with Large Diffeomorphic Deformations*, SIAM Journal on Imaging Sciences, 11 (2018), pp. 575–617.

[18] C. Chen and G. Xu, *Gradient-flow-based semi-implicit finite-element method and its convergence analysis for image reconstruction*, Inverse Problems, 28 (2012), p. 035006.

[19] C. Chen and G. Xu, *The Linearized Split Bregman Iterative Algorithm and Its Convergence Analysis for Robust Tomographic Image Reconstruction*, Tech. Report 13-66, UCLA CAM Reports, 2013.

[20] C. Chen and G. Xu, *A new linearized split Bregman iterative algorithm for image reconstruction in sparse-view X-ray computed tomography*, Comput. Math. Appl., 71 (2016), pp. 1537–1559.

[21] K. Chen and D. A. Lorenz, *Image sequence interpolation using optimal control*, Journal of Mathematical Imaging and Vision, 41 (2011), pp. 222–238.

[22] M. Dawood, F. Büther, X. Jiang, and K. P. Schäfers, *Respiratory motion correction in 3-D PET data with advanced optical flow algorithms*, IEEE Trans. Med. Imaging, 27 (2008), pp. 1164–1175.

[23] M. Dawood, F. Büther, N. Lang, O. Schober, and K. P. Schäfers, *Respiratory gating in positron emission tomography: a quantitative comparison of different gating schemes*, Med. Phys., 34 (2007), pp. 3067–3076.

[24] M. Dawood, N. Lang, X. Jiang, and K. P. Schäfers, *Lung motion correction on respiratory gated 3-d PET/CT images*, IEEE Trans. Med. Imaging, 25 (2006), pp. 1164–1175.

[25] M. Dawood, L. Stegger, X. Jiang, O. Schober, M. Schäfers, and K. P. Schäfers, *Optimal number of respiratory gates in positron emission tomography: a cardiac patient study*, Med. Phys., 36 (2009), pp. 1775–1784.

[26] P. Dupuis, U. Grenander, and M. Miller, *Variational Problems on Flows of Diffeomorphisms for Imaging Matching*, Quarterly of Applied Mathematics, 56 (1998), pp. 587–600.

[27] W. v. Elmpt, J. Hamill, J. Jones, D. De Ruysscher, P. Lambin, and M. Öllers, *Optimal gating compared to 3D and 4D PET reconstruction for characterization of lung tumors*, Eur. J. Nucl. Med. Mol. Imaging, 38 (2011), pp. 843–855.
[28] H. Gao, J. Cai, Z. Shen, and H. Zhao, *Robust principal component analysis-based four-dimensional computed tomography*, Phys. Med. Biol., 56 (2011), pp. 3181–3198.
[29] F. Gigengack, X. Jiang, M. Dawood, and K. Schäfers, *Motion Correction in Thoracic Position Emission Tomography*, Springer Briefs in Electrical and Computer Engineering, Springer, 2015.
[30] F. Gigengack, L. Ruthotto, M. Burger, C. H. Wolters, X. Jiang, and K. P. Schäfers, *Motion correction in dual gated cardiac PET using mass-preserving image registration*, IEEE Trans. Med. Imaging, 31 (2012), pp. 698–712.
[31] U. Grenander and M. Miller, *Pattern Theory. From Representation to Inference*, Oxford University Press, 2007.
[32] W. Grootjans, L.-F. de Geus-Oei, A. P. W. Meeuwis, S. Charlotte, M. Gotthardt, W. Oyen, and E. Visser, *Amplitude-based optimal respiratory gating in positron emission tomography in patients with primary lung cancer*, Eur. Radio., 24 (2014), pp. 3242–3250.
[33] J. Hinkle, M. Szegedi, B. Wang, B. Salter, and S. Joshi, *4D CT image reconstruction with diffeomorphic motion model*, Medical Image Analysis, 16 (2012), pp. 1307–1316.
[34] X. Jia, Y. Lou, B. Dong, Z. Tian, and S. Jiang, *4D computed tomography reconstruction from few-projection data via temporal non-local regularization*, in Medical Image Computing and Computer-Assisted Intervention – MICCAI 2010, T. Jiang, N. Navab, J. Pluim, and M. Viergever, eds., vol. 6361 of Lecture Notes in Computer Science, Springer-Verlag, 2010, pp. 143–150.
[35] N. Lang, M. Dawood, F. Büther, O. Schober, M. Schäfers, and K. P. Schäfers, *Organ movement reduction in PET/CT using dual-gated list-mode acquisition*, Med. Phys., 16 (2006), pp. 93–100.
[36] J. Liu, X. Zhang, H. Zhao, Y. Gao, D. Thomas, D. Low, and H. Gao, *5D respiratory motion model based image reconstruction algorithm for 4D cone-beam computed tomography*, Inverse Problems, 31 (2015), p. 115007.
[37] G. Lucignani, *Respiratory and cardiac motion correction with 4D PET imaging: shooting at moving targets*, Eur. J. Nucl. Med. Mol. Imaging, 36 (2009), pp. 315–319.
[38] M. I. Miller, A. Trouvé, and L. Younes, *On the metrics and Euler-Lagrange equations of computational anatomy*, Annual Reviews of Biomedical Engineering, 4 (2002), pp. 375–405.
[39] M. I. Miller, A. Trouvé, and L. Younes, *Geodesic shooting for computational anatomy*, Journal of Mathematical Imaging and Vision, 24 (2006), pp. 209–228.
[40] M. I. Miller, A. Trouvé, and L. Younes, *Hamiltonian systems and optimal control in computational anatomy: 100 years since D'arcy Thompson*, Annual Review of Biomedical Engineering, 17 (2015), pp. 447–509.
[41] F. Natterer, *The Mathematics of Computerized Tomography*, vol. 32 of Classics in Applied Mathematics, SIAM, 2001.
[42] H. Nien and J. A. Fessler, *Fast X-ray CT image reconstruction using a linearized augmented Lagrangian method with ordered subsets*, IEEE Trans. Med. Imag., 34 (2015), pp. 388–399.
[43] O. Öktem, C. Chen, N. O. Domaniç, P. Ravikumar, and C. Bajaj, *Shape based image reconstruction using linearised deformations*, Inverse Problems, 33 (2017), p. 035004.
[44] L. Ritschl, S. Sawall, M. Knaup, A. Hess, and M. Kachelriess, *Iterative 4D cardiac micro-CT image reconstruction using an adaptive spatio-temporal sparsity prior*, Phys. Med. Biol., 57 (2012), pp. 1517–1525.
[45] L. Rudin, S. Osher, and E. Fatemi, *Nonlinear total variation based noise removal algorithms*, Phys. D, 60 (1992), pp. 259–268.
[46] O. Scherzer, M. Grasmair, H. Grossauer, M. Haltmeier, and F. Lenzen, *Variational Methods in Imaging*, vol. 167 of Applied Mathematical Sciences, Springer-Verlag, New York, 2009.
[47] H. Schumacher, J. Modersitzki, and B. Fischer, *Combined reconstruction and motion correction in SPECT imaging*, IEEE Trans. Nucl. Sci., 56 (2009), pp. 73–80.
[48] A. J. Schwarz and M. O. Leach, *Implications of respiratory motion for the quantification of 2D MR spectroscopic imaging data in the abdomen*, Phys. Med. Biol., 45 (2000), pp. 2105–2116.
[49] E. Sidky, J. Jørgensen, and X. Pan, *Convex optimization problem prototyping for image reconstruction in computed tomography with the Chambolle–Pock algorithm*, Physics in medicine and biology, 57

(2012), p. 3065.
[50] E. Sidky, C. Kao, and X. Pan, *Accurate image reconstruction from few-views and limited-angle data in divergent-beam CT*, Journal of X-ray Science and Technology, 14 (2006), pp. 119–139.
[51] K. Taguchi and H. Kudo, *Motion Compensated Fan-Beam Reconstruction for Nonrigid Transformation*, IEEE Trans. Med. Imaging, 27 (2008), pp. 907–917.
[52] Q. Tang, J. Cammin, S. Srivastava, and K. Taguchi, *A fully four-dimensional, iterative motion estimation and compensation method for cardiac CT*, Med. Phys., 39 (2012), pp. 4291–4305.
[53] D. Thompson, *On Growth and Form*, Cambridge University Press, New York, 1945.
[54] A. Trouvé, *Diffeomorphisms Groups and Pattern Matching in Imaging Analysis*, International Journal of Computer Vision, 28 (1998), pp. 213–221.
[55] A. Trouvé and L. Younes, *Shape spaces*, in Handbook of Mathematical Methods in Imaging, S. Otmar, ed., Springer-Verlag, 2015, pp. 1759–1817.
[56] A. Van Der Gucht, B. Serrano, F. Hugonnet, B. Paulmier, N. Garnier, and M. Faraggi, *Impact of a new respiratory amplitude-based gating technique in evaluation of upper abdominal PET lesions*, Eur. J. Radiol., 83 (2014), pp. 509–515.
[57] F.-X. Vialard, L. Risser, D. Rueckert, and C. J. Cotter, *Diffeomorphic 3D image registration via geodesic shooting using an efficient adjoint calculation*, International Journal of Computer Vision, 97 (2012), pp. 229–241.
[58] Y. Wang, E. Vidan, and G. W. Bergman, *Cardiac motion of coronary arteries: variability in the rest period and implications for coronary MR angiography*, Radiology, 213 (1999), pp. 751–758.
[59] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, *Image quality assessment: From error visibility to structural similarity*, IEEE Trans. Image Process., 13 (2004), pp. 600–612.
[60] L. Younes, *Shapes and Diffeomorphisms*, vol. 171 of Applied Mathematical Sciences, Springer-Verlag, 2010.